Spring 2021

Analyzing Volatility and Policy Changes in the Financial Market: Three Essays in Applied Finance

Rebecca Anne Bommarito
Claremont Graduate University

Analyzing Volatility and Policy Changes in the Financial Market: Three Essays in Applied Finance

By
Rebecca Bommarito

Claremont Graduate University
2021

Approval of the Dissertation Committee

This dissertation has been duly read, reviewed, and critiqued by the Committee listed below, which hereby approves the manuscript of Rebecca Bommarito as fulfilling the scope and quality requirements for meriting the degree of Doctor of Philosophy in Economics.

Pierangelo DePace, Chair
Claremont Graduate University and
Pomona College
Associate Professor of Economics

Tom Willet
Claremont Graduate University and
Claremont McKenna College
Director, Claremont Institute for Economic Policy Studies
Horton Professor of Economics

Clemens Kownatzki
Pepperdine Graziadio Business School
Department Chair of Accounting, Finance & Real Estate
Assistant Professor of Finance

Hisam Sabouni
Claremont Graduate University
Clairvoyant Financial

Abstract

Analyzing Volatility and Policy Changes in the Financial Market:
Three Essays in Applied Finance

By

Rebecca Bommarito

Claremont Graduate University: 2021

Volatility in financial markets make forecasting, or in other words estimating what will happen in the future, a difficult task. Too often forecasts are made but hardly ever revisited to see how accurate the forecast was and if not, why? The three chapters of my dissertation are focused on examining volatility in financial markets from changes in investors' trading behavior as well as studying the characteristics of forecast error of various financial securities. Often, the accuracy of these forecasts rely on the estimates made for future volatility.

In my first chapter, we[1] analyze the predictive power of the Black-Scholes-Merton (BSM) model on a data set of options on the SPDR SP 500 Trust ETF (SPY). We leverage the full options chain to analyze the full forecasted distribution of prices through N(d2), which we compare to the distribution of prices of SPY. Using non-parametric GOF tests, such as the Kolmogorov–Smirnov and Anderson-Darling tests, we are able to analyze whether two different distributions come from the same underlying population distribution. We find that BSM tends to overestimate the tails in the implied probability distribution when further away in expiration, compared to the empirical price path of SPY. The resulting comparison gives way to visualizing and testing the ability for the BSM to predict the likelihood of options expiring in-the-money. Our findings suggest the BSM, in most cases, correctly estimates the underlying risk adjusted probabilities only a few days out from expiration, which may be attributed to the uncertainty in traders to foresee market movements until an option is close to expiration. However, this behavior is more pronounced during crisis periods, where the BSM tends to correctly estimate the likelihood of tail events occurring more often than during periods of market normalcy.

In my second chapter, my co-authors and I[2] study the characteristics of error in economic forecasts over time. We focus on explaining the variation errors of the survey of professional economic forecasts (SPF) across three financial securities by isolating the effects of changes in fiscal and monetary policy as well as changes in various macroeconomic indicators. We

[1]This is joint work with Nasser Khalil, Clemens Kownatzki and Hisam Sabouni

[2]This is joint work with Hisam Sabouni

examine if it is changes in government policy or changes in macroeconomic indicators (or market conditions) that is primarily responsible for increases in SPF forecast error. We use a principal component analysis to first perform orthogonal dimension reduction of our macroeconomic indicator variables and use the first two principal components as an overall measure for market conditions. We then use a linear regression to test whether market conditions or monetary and/or fiscal policy is primarily responsible for increases in SPF forecast error of three securities' yields: the three month Treasury bill, Moody's AAA corporate bond and the Ten year Treasury bill. We find increases in monetary policy via the EFFR affects the short-term security in our analysis to a large magnitude, but increases in overall market conditions affect all securities in our analysis to a smaller but significant degree.

In my third chapter, I explore an anomaly that exists in the U.S. equity market that has not been documented before; investors' reactions to earnings announcements are not only asymmetric, but seasonal. Knowing which months experience larger variation than others, investors may incorporate financial derivatives such an options to hedge downside risk. Using a fixed effects linear regression, I first examine the effect an earnings beat and earnings miss have on abnormal returns; which are calculated by a CAPM-GARCH model. I find an earnings beat on average has large significant increases in firms' abnormal returns while an earnings miss, or a negative earnings surprise, has limited downside impacts. Examining this effect further at the month level, I find investor's reactions are extremely large to earnings beats announced in months June and to earnings misses announced in December compared to other months.

Acknowledgments

I would like to start by first thanking the faculty and administration of Claremont Graduate University (CGU) for their continued support and encouragement throughout my research studies and experience in the doctoral program. I feel extremely grateful for this opportunity and to have met the professors and colleagues that I did along the way, for this would not have been possible without them.

I am very thankful to my two academic advisors, Greg DeAngelo and Hisam Sabouni, for guiding my research ideas and providing a collaborative environment. Greg brought me into the Computational Justice Lab (CJL) during my second year of CGU and introduced me to several other members on campus who could assist me in my research endeavors within applied finance. Another member of the CJL lab, Hisam Sabouni, was really the catalyst for my interest in applied finance and one of my main motivators during the doctoral program. Hisam is a co-author on two of my dissertation chapters and member of my dissertation committee. I view Hisam as my mentor and am truly grateful for all the opportunities he has provided for me and time he has spent to improve my work.

Another huge motivator of my research studies as well as the chair of my dissertation committee is Pierangelo DePace. Pierangelo was incredibly supportive and would frequently meet with me to discuss research, the Ph.D. process, or anything else going on in my life. I developed by understanding of econometrics largely through Pierangelo's classes and feel extremely blessed to have had such a thoughtful and thorough professor.

I am grateful to other members of my committee, Tom Willet and Clemens Kownatzki who both provided invaluable comments to the papers in my dissertation. Tom has been a pioneer in the areas of International Money & Finance and Public Policy and is a well respected figure across our community. His brilliant insights and thought provoking questions have truly elevated each of my research papers. I was introduced to Clemens through Hisam and he quickly became someone I would turn to for advice on my papers. Clemens helped edit and one of the papers in my dissertation as is a coauthor on a second, which was a wonderful collaborative experience for me.

I would also like to acknowledge two faculty members from my undergraduate studies, Brian Jenkins and Kim Makuch, whose continued support and guidance helped lead me to where I am today. Brian was one of my economic lectures and hired me as an economic tutor. It was through Brian's classes that I developed a passion for economics and the desire to continue my education in economics at the graduate level. Kim was a TA of mine and someone I developed a professional and close personal relationship with. At the time

Kim was towards the end of finishing her Ph.D. and guided me through the entire process of applying to grad school, edited papers for my applications and kept it real about the challenges I would face if admitted. I am beyond grateful to have had such a role model and friend during that process and stage of my life.

I want to thank another coauthor, Nasser Khalil, who was a pleasure to work with on one of the papers in my dissertation. Nasser and I had a very fluid writing style while working together and complemented each others skill sets quite which helped to elevate our paper.

Two members of the CJL lab and close personal friends I would also like to thank are Maryah Garner and Minjae Yun. I first met Maryah as a TA for one of my classes and over the years has become my best friend and big sister in the graduate program. Countless times Maryah would read over my papers and edit them, discuss questions I had from class or research, or hype me up before a big test. Minjae was the first person I met at CGU. We bonded instantaneously and eventually became roommates. We did everything together and it really made the graduate process a lot more fun. I truly do not know how I would have gotten through all those classes, tests and qualifying exams without Minjae and her hilarious nature.

Most of all, I would like to thank my significant other Keenan. I appreciate how Keenan has believed in me throughout every step of the Ph.D. process. From long exhausting days to anxiety inducing tests, he was always by my side. Keenan put my dreams above his own to support me on this journey, and I am forever grateful. Lastly, I would like to thank both of my parents for believing in every dream of mine and encouraging me throughout my graduate studies.

Contents

1 Chapter 1: Analyzing the Predictive Power of Black-Scholes

Coauthored with Nasser Khalil, Clemens Kownatzki and Hisam Sabouni

1.1 Introduction

For some time, economists have utilized options pricing models, such as the Black-Scholes Merton (BSM) model, to not only estimate an option's fair value, but to forecast the volatility of the underlying asset derived from the implied volatility in options. One way to examine this is to back out of BSM what the option is "implying" about the underlying future expected volatility, also known as "implied volatility". The price of an option is dependent on the breadth of possible scenarios that may occur over the life of an option. Much of the attraction of the BSM model has been on modeling the risk and sensitivity of the options price of the underlying security. A natural product of the BSM, however, is the underlying risk-adjusted probabilities used to price options. With an options chain, probabilities are assigned as to the likelihood the underlying asset will be above the strike price at expiration. Little, if any research, has focused on the implications of risk-adjusted probabilities derived by the BSM model and the weight given to "deep-in' and "out of' the money contracts. Our analysis aims to address the insights provided by the market expectations of risk adjusted probabilities as backed out by the BSM, and the ability to forecast whether an option will expire in the-money.

During market crashes and large price swings, increases in volatility tend to suggest a higher likelihood of tail events occurring. Canonical models, such as that of Merton (1973) and Black and Scholes (1973) highlight the relative importance of volatility in determining the price of an option. In this paper, we seek to examine the forecasting abilities of Black-Scholes, and examine how fast the implied distribution of BSM adjusts to changes in prices and thus affects the accuracy of the forecast. As individuals' perceptions of volatility change over time, the price of options can vary dramatically. Kownatzki and Sabouni (2019) show options have an extra insurance premium and thus options prices are inflated relative to what we would observe based on historical volatility of the underlying assets. Put options have slightly higher levels of implied volatility because prices fall faster than they rise[3], resulting in a higher demand for put options to hedge downside risk. Professional option

[3]Kownatzki and Sabouni (2019) present the case of option an options strangle strategy suffering from deep losses from a maximum drawdown during crisis periods.

traders, as a result, may be aware that option prices are inflated relative to what is observed in the market, resulting in the net selling of options. This notion suggests that in periods of market normalcy, options tend to overestimate the likelihood of exercise at expiration. In most cases, implied volatility is highest in options with strike prices further away from the current market price. Furthermore, options expire and become worthless if they are out-of-the-money. Typically, longer-dated options have lower levels of implied volatility when short-dated volatility is low. Our analysis focuses on documenting how the BSM model assigns risk adjusted probabilities to these phenomena.

Deriving the implied volatility from BSM can provide significant benefits to investors, such as providing an estimate of the future variability for the assets underlying their options contract. Implied volatility provides a more robust estimate of risk compared to using historical data. Option sensitivity to changes in implied volatility, or vega[4], for at the-money option prices is greater than in the-money or out-the-money options Poon and Granger (2003). Therefore, implied volatility estimates are derived from strike prices for ITM, ATM, and OTM options on heavily traded securities, and is used as an alternative measure of risk compared to historical volatility. Alternatively, research has centered on observing the VIX (S&P 500 Volatility Index) as an overall barometer of market risk using a weighted average of implied volatility.

Much of the previous literature surrounding implied volatility from BSM utilize point estimates, which acts as an estimate of expected future volatility of an options contract Fleming (1998). In this analysis, we model the reverse cumulative distribution function (CDF) of N(d2) 60/90/120/180 days[5] from expiration for call and puts separately. We track each and every option traded on the SPY every day from each of our respective start periods until the expiration of the options. A novel contribution of this paper is that we utilize the full options chain to analyze the full forecasted distribution (density forecast) of prices through $N(d_2)$ from the initial transaction date all the way to expiration. Until recently, most forecasts were provided as point estimates, with measures of uncertainty, such as standard errors, included. Recent trends include forecasters providing density or interval forecasts along with their point forecasts. Density forecasts are more universal than point forecasts as they provide information on the full forecasted distribution of a random variable. The availability of density forecasts allow users to focus on specific moments of the distribution that are of interest (such as the predicted mean, median, quantile, etc.), depending on the

[4]See Figure 11 for a detailed look at the behavior of vega across various degrees of moneyness.

[5]Here, 120 and 180 DTE are double the amount of 60 to 90, respectively. Therefore, 120 and 180 DTE provide double the amount of information and data needed to analyze various sub-periods in our sample.

user's specific loss function by which the forecast is evaluated. Since density forecasts contain all probabilistic information about the random variable, every user's needs will be satisfied, regardless of the loss function[6]. Our paper analyzes the evolution of the distribution of $N(d_2)$ over time, which demonstrates the change in how investors implicitly assign probabilities to the likelihood that an option will expire in the-money. This dynamic is represented by a change in the shape of the curve as an options chain approaches maturity. By observing the similarity between the underlying price path and implied probability distribution, investors are better able to assess which possible strike prices are more likely to fall in-the-money at expiration. Closer to expiration, options that are closest to the underlying spot price are assigned greater probabilities of being profitable. The uncertainty in the likelihood of an option expiring in the-money is partly demonstrated by the same behavior exhibited in the volatility smile. Farther from expiration, there is considerable uncertainty in deep in and out-the money options. Our interest is in observing this shift in uncertainty during market crashes or black swan events, where uncertainty is at its highest [7]. Rather than focusing on risk measures such as implied volatility, we utilize these events in the context of changing risk-adjusted probabilities of $N(d_2)$. The probability measure of $N(d_2)$ provides us with a forecast of what the underlying price distribution may look like at expiration, rather than observing a point estimate of implied volatility.

By observing unique expiration dates in a sample of data, we are able to observe and compare how risk-adjusted probabilities are assigned as to the likelihood of an option expiring in the money. We estimate the cumulative distribution of an options chain across time prior to expiration to analyze how well the distribution predicts the underlying price path of an underlying security. Increasingly, comparisons not only include the means of two samples, but entire distributions. Aıt-Sahalia et al. (2001) apply a similar study to S&P 500 index to compare the distribution of risk neutral probability measures on the underlying index returns, but find that the two distributions differ substantially. Our approach differs from Aıt-Sahalia et al. (2001) in that we observe probability measures, implicitly, from the BSM, rather than a non-parametric model of options prices. It is then possible to ask whether the two samples have identical or different distributions, and whether the distributions differ at the median (or specified quantile) and/or whether the CDFs differ at a particular value.

[6]In a specific case, if the user's loss function depends not only on the point forecast, but on a two-sided prediction-interval, or even the entire density (i.e., the loss function is a scoring-rules), a density forecast is favorable to a point forecast.

[7]Peak volatility in the market can also be seen by the behavior of the VIX, which is the implied risk of the overall market (S&P 500 Index).

One way to compare two distributions is to use a Kolmogorov-Smirnov (K-S) (1939) two sample non-parametric GOF test, which tests if two different distributions come from the same underlying population distribution.

Another test similar to the K-S test is the Anderson–Darling (A-D) (1952) Scholz and Stephens (1987) two sample non-parametric GOF test. These two tests differ only slightly in where the majority of the power lies; the center or the tails of the distribution. The K-S test tends to be more sensitive near the center of the distribution, while the A-D test tends to be more sensitive near the tails. Comparing the empirical CDF of the price path relative to the predicted risk-adjusted probabilities allows for a quantitative analysis of how similar both probability curves evolve over time. The proceeding paper addresses the comparison between the risk-adjusted probability measures provided by the BSM model, and an empirical CDF of historical prices on the underlying option.

Since theta decay accelerates exponentially as we approach expiration, this then causes the distribution of risk adjusted probabilities to collapse towards a narrower set of possible price paths. We show that this holds true not only for the implied BSM distribution compared to the actual distribution, but also when comparing the implied distribution of BSM to a Monte Carlo simulation of implied probability estimates. While the Monte Carlo randomly iterates/changes as the option gets closer to expiration, it more or less stays similar in shape (an S curve) to the BSM implied distribution; both far and close to expiration. Overall, the Monte Carlo simulation is closer to the true distribution when it is further away from expiration. Our analyses also observe periods in which the BSM is able to correctly predict the underlying price path of SPY. We contribute to the existing literature surrounding options by observing the full distribution of option chains across time, and assessing the forecasting ability of the market expectations backed out from the BSM in determining risk adjusted probabilities.

1.2 Methods

The contribution of Black and Scholes (1973) and Merton (1973) and their development of the BSM model provides an analytical framework for observing the evolution of European[8] option prices. Inputs from BSM provide useful insights into the behavior of options, and the underlying asset associated with the option. The BSM model assumes the option is

[8] Although we are using American options for our analysis on SPY, the differences between the two option types stems from the options price and ability of early exercise in American options. For a further discussion, see the **Appendix**.

European and can only be exercised at expiration, no dividends are paid out during the life of the option, no transaction costs, markets are efficient (i.e., market movements cannot be predicted) and assumes constant values for risk free rate of return. Lastly, BSM model assumes the price of the underlying asset follows a geometric Brownian motion with constant drift and volatility over the option duration. We relax the assumption of the traditional BSM framework in assuming, in the case of SPY, that dividends are non-zero. The following modified equations provide the underlying BSM model, adjusting for the addition of dividends in the model:

$$Call = S_0 e^{-qT} N(d_1) - K e^{-rT} N(d_2) \tag{1}$$

$$Put = K e^{-rT} N(-d_2) - S_0 N(-d_1) e^{-qT} \tag{2}$$

$$d_1 = \frac{ln(\frac{S_0}{K}) + (r_f - q + \frac{\sigma^2}{2})T}{\sigma\sqrt{T}} \tag{3}$$

$$d_2 = d_1 - \sigma\sqrt{T} \tag{4}$$

where, S_0 is the spot price of the underlying security, q is the continuously compounded dividend yield, $N(\cdot)$ represents the cumulative normal distribution, K is the strike price of the option, r is the risk-free rate, σ is the volatility of the underlying security, and T is the time to expiration expressed in years. The first/second term $S_t N(d_1)$ for a call/put option is the product of the discounted expected value of the stock price at maturity; conditional that $S > K$ or $S < K$ at maturity for a given $Prob(S > K)$ or $Prob(S < K)$ at maturity. $N(d_1)$ assumes that the option will expire in-the-money and then determines the probability of a particular in-the-money stock price. The second/first term, $Ke^{-rT}N(d_2)$, is the discounted exercise price times the probability the terminal stock price exceeds the exercise price. $N(d_2)$ equals the probability that the call/put will finish in the money in a risk-free world.

A main drawback of the BSM is the assumption of constant volatility across an options chain. This assumption is not reflected in the real world where different strike prices have different levels of implied volatility reflecting the fact that investors and traders assign higher premiums for options that allow them to protect their portfolios. Although BSM assumes a log-normal distribution (Gaussian Distribution) of price changes for the underlying asset, prices can have a larger skewness and kurtosis; meaning that high risk downward moves occur more often in the options market than a Gaussian distribution predicts. Due to the assumptions of a log-normal distribution on the underlying asset prices, implied volatility is quite similar across strikes according to BSM [9] . While the 1987 stock market crash had little change in observable macroeconomic fundamentals, market prices fell 20-25% and interest rates dropped about 1-2%. Due to this, market makers changed their assumptions about volatility being equally spread between out/in/at the-money options, which led to a shift in the prices market makers were willing to deal at; triggering a permanent shift in index option prices. Since the 1987 market crash, implied volatility for at the-money options have been lower than those far out the-money or far in the-money options. The Black-Scholes formula has been significantly under pricing short-maturity, deep out of the-money S&P 500 put options, revealing implied volatility values form a convex curve along the strike prices (Rubinstein (1994), Bates (2000)). The reason for this occurrence is that the market prices in a higher likelihood of a sharp downward price movement than an increase. It also means put options have higher implied volatility than call options since prices fall faster than they rise. This has led to the presence of the volatility skew. When the implied volatility for options with the same expiration date are mapped out on a graph, a smile or skew shape can be seen. Another way to view this is to observe for low strike prices, the underlying spot price is already in the-money for a given options chain. The probability that the option will expire in the-money at expiration is fairly high for deep in the-money contracts, and significantly smaller for deep out-the-money contracts. It is also possible to have reverse volatility smile if the implied volatility is higher on lower options strikes relative to the current market price. These are most commonly seen in index options or other longer-term options and occurs at times when investors are uncertain about the future and purchase puts to hedge their position.

Options are priced with expected future volatility. Instead of backing out implied volatility from the BSM, we instead calibrate the BSM using inputs, provided by our data set, to find $N(d_2)$. We first start by calculating the probability of exercise by using data on delta, which can be expressed for call[10] options as:

[9] The same assumptions don't hold for implied volatility on strikes farther from the underlying market price. Instead, there is a non-linear increase in implied volatility.

[10] Whereas for puts, delta is calculated as $N(d_1) - 1$.

$$\Delta = N(d_1) \tag{5}$$

We apply[11] the inverse of the normal distribution to Equation 5, and are left with a value for d_1. Hence, we are then able to apply Equation 4 to calculate $N(d_2)$. For each trading day and unique options expiration, we apply this procedure for a complete options chain. Likewise, for each day and expiration, an implied cumulative distribution of probabilities is formed across all strike prices. Given that this is a probabilistic forecast, most of the probability density remains close to the strike price, which looks similar to a bell curve. The same process is repeated for every sample options chain dating back 60/90/120/180 days from expiration, which then allows us to observe any possible patterns that may arise for an options chain of calls or puts at for a given expiration date. The focus of the paper is to observe the risk-adjusted probability measure of $N(d_2)$ and $N(-d_2)$ for call and put options, respectively, of an option being exercised. Across an options chain, we are able to construct, from the BSM framework, a cumulative distribution function of $N(d_2)$ (or $N(-d_2)$). Using implied probability[12] and number of days until maturity, we apply equation 4 to calculate d_2 and market expectations of probability measure $N(d_2)$.

As described earlier, the Kolmogorov-Smirnov [13] (K-S) (1939) is a two sample non-parametric GOF test, which tests if two different distributions come from the same underlying population distribution:

$$D_k(F, G) = \sup_t |F(t) - G(t)| \tag{6}$$

Equation 6 represents the two-sample KS test statistic, where $F(t)$ and $G(t)$ are two sample distributions to test for the null hypothesis of $F(t) = G(t)$. In our analysis, we use the cumulative distribution of $N(d_2)$ ($N(-d_2)$ for puts) to compare to the empirical distribution of realized stock prices during the forecasted option period.

The K-S test is based on the empirical distribution function (ECDF). Given N ordered data points $Y_1, Y_2, ..., Y_N$, the ECDF is defined as

[11] Although Δ of a call is equal to $N(d_1)$ in the case of a non-dividend paying stock, we assume the difference in calculating delta for dividend paying equities to be negligible. For the purposes of our study, Equation 5 holds for low dividend yielding equities such as SPY.

[12] We note here that implied volatility values are provided in the data set from IVolatility.com, which utilizes a 100 step binomial tree to create these estimates. For a further explanation, see the **Appendix**.

[13] See Lehmann and Romano (2006) for a further look at the application of the KS test for one or two sample comparisons.

$$E_N = n(i)/N \tag{7}$$

where n(i) is the number of points less than Y_i and the Y_i are ordered from smallest to largest value. This is a step function that increases by 1/N at the value of each ordered data point.

An attractive feature of this test is that the distribution of the K-S test statistic itself does not depend on the underlying cumulative distribution function being tested; it is non-parametric. A second advantage is that it is an exact test; it does not depend on an adequate sample size for the estimates to be valid. Despite these positive features, the K-S test does suffer from a few limitations. The K-S test tends to be more sensitive near the center of the distribution than at the tails. The variance of the sample CDF in the tails is smaller than near the median. That is, it is "harder" to achieve the critical value of D in the tail region than in the middle, so the test generally finds deviations more toward the middle than right up at the ends.

An alternative to the K-S test is the Anderson-Darling (A-D) (1952) (Stephens 1974) test for normality which gives more weight to the tails than the K-S test and also uses information from all of the differences, not just the largest one. An extension of the A-D test comes in the form of a K-sample test, which can be used to test whether several collections of random distributions can be modeled as coming from the similar continuous populations. Our analysis applies a 2-sample A-D test, which can be denoted by

$$A^2_{nm} = \frac{n \cdot m}{N} \int_{-\infty}^{\infty} \frac{(F_n(x) - G_m(x))^2}{H_N(x) \cdot (1 - H_N(x))} dH_N(x) \tag{8}$$

where n and m are the sample sizes of the two random distributions of $F_n(x)$ and $G_m(x)$, N is the combined sample size equal to $n + m$, and $H_n(x) = \frac{(nF_n(x)+G_m(x))}{N}$. The A-D test takes the weighted average of each sample observation, and takes the squared sum difference between the two distributions. This approach differs significantly from the K-S test, where the largest point wise difference is used as a critical value for comparing distributions. Similarly to the K-S test, the K-sample A-D test is non-parametric goodness of fit measure. The A-D test makes use of the specific distribution in calculating critical values, which has the advantage of allowing a more sensitive test, but the disadvantage that critical values must be calculated for each distribution. For robustness, we use both the K-S statistic and A-D statistic to compare the empirical distribution of a given options chain to the actual.

The K-S and A-D test are the main GOF tests we apply to compare distributions of risk-adjusted probabilities from the BSM, and empirical CDF of SPY price path. Such an approach is instrumental in assessing a goodness of fit measure to our analysis, and the predictive power of the BSM. The risk-adjusted probabilities from the BSM forecasts the likelihood an option will expire in the-money, at each point in time prior to expiration. That is, prior to expiration, estimates how well the BSM predicts the empirical distribution of SPY prior to expiration. The K-S test is first applied to a sample expiration date on a call and put options chains, separately, for SPY. We proceed to test the similarity in distributions of $N(d_2)$ from the BSM at each point in time, 180/120/90/60 days prior to expiration. From the 180/120/90/60 day sample period prior to expiration, we create an empirical CDF (ECDF) using the spot price of SPY. For each point in time, all possible strike prices of the options chain act as inputs into the empirical distribution function. The same process can be repeated for each unique option expiration date.

1.3 Data

From the period of January 2005 to December 2020, data is obtained from iVolatility on the SPDR S&P 500 ETF Trust (SPY)[14]. Data is obtained for American[15] style options on SPY as a baseline measure for the analysis, which includes the underlying closing spot price, estimates of the Greeks, implied volatility, volume and open interest of traded option contracts, and the number of days until expiration. We outline the Greeks for call and put options in Table 12 by calculating the mean across a range of moneyness. Table 10 and 11 summarize the average implied volatility differential, which we calculate as the difference of implied volatility from the closest at-the-money contract in an options chain. For deep out-the-money puts and deep in-the-money calls, the implied volatility differential peaks during recessionary periods.

The use of SPY options contracts is an ideal barometer of market activity, and has a high degree of open interest, variety of strike prices and maturities[16]. Tables 1 and 15 provide an average measure of activity in SPY options through open interest, and liquidity measured by the bid-ask spread. Open interest represents the number of contracts that have yet to be exercised or offset. Shorter dated options, up to 90 days to expiration, have a greater

[14] We chose 2005 as a start date given that there is very little options data available prior to this.

[15] Although BSM assumptions rely on the ability not to exercise the option early, American options can still be approximated by this framework. For a further discussion, see the **Appendix**.

[16] Note: a substantial amount of time in our sample is plagued by artificially low interest rates as a result of the monetary stimulus following the financial crises in 2008.

Table 1: Descriptive Statistics for SPY Call Options

	Range of Days to Expiration			
	0 to 60	60 to 90	90 to 120	120 to 180
Mean Length of Strikes	77	85	86	83
Unique Number of Contracts	78,805	21,063	17,483	13,787
Mean Open Interest by K/S_0				
Less than 0.85	618	497	592	581
0.85 to 0.95	2,682	2,747	2,445	2,337
0.95 to 1.05	10,508	11,650	7,582	6,428
1.05 to 1.15	9,165	10,989	8,105	7,359
Greater than 1.15	5,029	4,222	3,612	3,176
Mean Bid-Ask Spread by K/S_0				
Less than 0.85	0.3681	0.3596	0.4087	0.4119
0.85 to 0.95	0.2855	0.1869	0.2057	0.2313
0.95 to 1.05	0.0654	0.0682	0.0906	0.1238
1.05 to 1.15	0.0536	0.0338	0.0579	0.0862
Greater than 1.15	0.1625	0.0305	0.0566	0.0605

Note: We introduce summary statistics on the sampled data for SPY call options contracts. The first descriptive stat shows the average number of strike prices within an options chains. We also count the number of unique contracts within the sample, which is shown in the second row. The remaining rows illustrate liquidity and market activity in call options contracts, by taking the average open interest and bid-ask spread for a given range of moneyness.

number of long positions across a range of moneyness, which is also evident by the number of outstanding contracts 60 to days to expiration. Note, however, that the number of strike prices in an options chain decreases towards a shorter range of days of expiration, suggesting deeper in the-money and out the-money have fewer long positions in the underlying asset. As an option approaches expiration, more contracts are created for strike prices that were previously not offered in an options chain. Increases in the bid-ask spread for both SPY call and put options are reflective of those options moving deeper in-the-money, and where less active trading occurs.

Our analysis excludes observations for options that have zero open interest, as no active

Table 2: Descriptive Statistics for SPY Put Options

	Range of Days to Expiration			
	0 to 60	60 to 90	90 to 120	120 to 180
Mean Length of Strikes	86	100	99	93
Unique Number of Contracts	87,915	24,838	20,049	15,491
Mean Open Interest by K/S_0				
Less than 0.85	8,272	6,962	6,373	6,236
0.85 to 0.95	17,775	17,367	11,963	10,632
0.95 to 1.05	13,196	12,764	7,946	6,388
1.05 to 1.15	3,629	2,492	2,390	2,093
Greater than 1.15	1,875	1,333	1,178	915
Mean Bid-Ask Spread by K/S_0				
Less than 0.85	0.016	0.023	0.030	0.045
0.85 to 0.95	0.019	0.035	0.046	0.071
0.95 to 1.05	0.080	0.078	0.092	0.120
1.05 to 1.15	0.437	0.302	0.337	0.330
Greater than 1.15	0.702	0.503	0.592	0.638

Note: We introduce summary statistics on the sampled data for SPY put options contracts. The first descriptive stat shows the average number of strike prices within an options chains. We also count the number of unique contracts within the sample, which is shown in the second row. The remaining rows illustrate liquidity and market activity in put options contracts, by taking the average open interest and bid-ask spread for a given range of moneyness.

trading occurs in these contracts. SPY options contracts are initialized many days [17] out from expiration, and allows for a more expansive data set to observe the changes to the shape of our estimated probability distribution. We implement the analysis of distributional similarity for a range of options contracts with days to expiration between 0 to 180 days, and hence exclude contracts with longer expiration dates. A distinction is made here with regards to how 'days to expiration' is denoted in our sample; data provided by iVolatility measures days to expiration in terms of calendar days, rather than trading days. We apply

[17]Expiration can occur well over a year from when an options contract or chain is initialized.

Equation 9 to calculate an approximate[18] length of trading days based on calendar days of 60/90/120/180, where Int takes the closest integer of the inner bracket. To accommodate for market closures during holidays, we subset for options chains with expiration lengths of at least 40/60/80/120 trading days for a sample of 60/90/120/180 calendar days.

$$Number\ of\ Trading\ Days \approx Calendar\ Days - Int\left[\frac{Calendar\ Days}{7}\right] * 2 \tag{9}$$

The filtered data is used to generate the full probability distribution of price ranges dating back an n number of days, observing the empirical distribution to the actual for a given options chain. Our sample consists of 14,466,550 observations, 203,474 unique contracts, and across 1,028 unique option expiration dates for both put and call option chains. For our analysis, we observe options chains 60/90/120/180 calendar days out from expiration to observe the distribution of probability measure $N(d_2)$. The theoretical distribution of $N(d_2)$ provided by the observed option chains on either call or puts, at a given point in time prior to expiration, can be compared to the empirical distribution of the underlying SPY closing price.

As an example, we take an initial look at the observed prices for a sample options chain which expires on 2012-11-16. Figure 1 displays the empirical CDF of the price path relative to the predicted risk-adjusted probabilities as of 2012-10-05, 42 days before expiration. Through $N(d_2)$ we back out the red and blue lines for call and put options. The red and blue lines are a snapshot in time. The green line is the observed time period of the actual price path; it shows how far actual prices have moved in the last 42 days of this particular option chain's life. We aim to compare how well the market expectations as backed out by the BSM is able to predict the underlying price path distribution of SPY (green line of Figure 1) using the estimated values of $N(d_2)$ of an options chain. As we get closer to expiration, the time value in option prices decreases exponentially; also known as theta decay. This is because there is less time/less probability of the option to end up in the-money at expiration. From Figure 1, it is possible to see that options are pricing in outlier events too much. That is, BSM is assigning larger probability density to tails to strike prices which are very unlikely to occur in the market. For example, with S_t = 145 and K = 80, implied cumulative Prob($S_t > K$) $\approx$ 100%, actual cumulative Prob($S_t > K$) $\approx$ 92%. We expand upon this analysis by not only observing the empirical CDF to the actual CDF on a random day before expiration,

[18]This formula does not fully account for all holidays within a given period of 60/90/120/180 days, when the market may typically be closed.

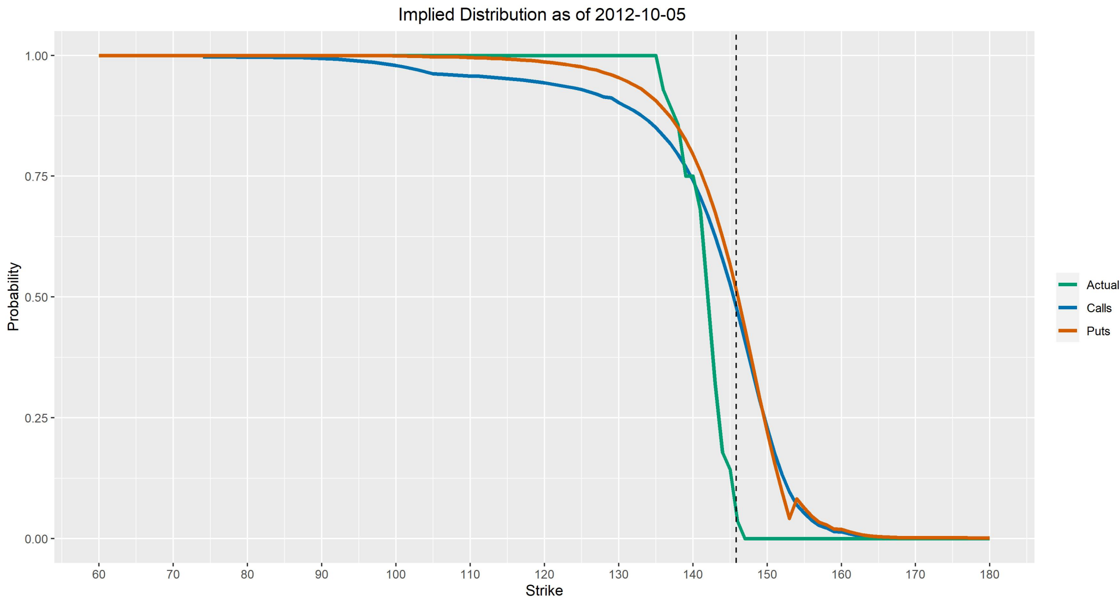


Figure 1: **BSM Implied CDF versus Actual CDF**

Note: The red line is the empirical CDF of the price path for put options and the blue line is the empirical CDF of the price path for call options. The green line is the actual price path for a randomly sampled options chain which expires on 2012-11-16. Therefore, the red and blue lines are the the predicted risk-adjusted probabilities as of 2012-10-05, 42 days prior to expiration. For each strike price, a corresponding point along the curves denotes the estimated probability as of the time of measurement of the security surpassing the strike price at the time of expiration.

but by observing these two distributions 60/90/120/180 days to expiration and every day after until the option expires.

Figure 2 presents a sample 3D plot of the observed distribution of $N(d_2)$ from BSM across 180 days for a call options chain that expires on 2005-12-16. The plot shows how the distribution becomes more vertical closer to maturity. Closer to maturity, the possible price path of where the option will likely expire becomes more apparent. Early in the life of the option, we see a greater spread of risk adjusted probability measures across strike prices. Deeper out of the money contracts have lower probability mass compared to options with strike prices at or near the the underlying spot price. This suggests BSM is slow in adjusting the predicted risk-adjusted probabilities for SPY.

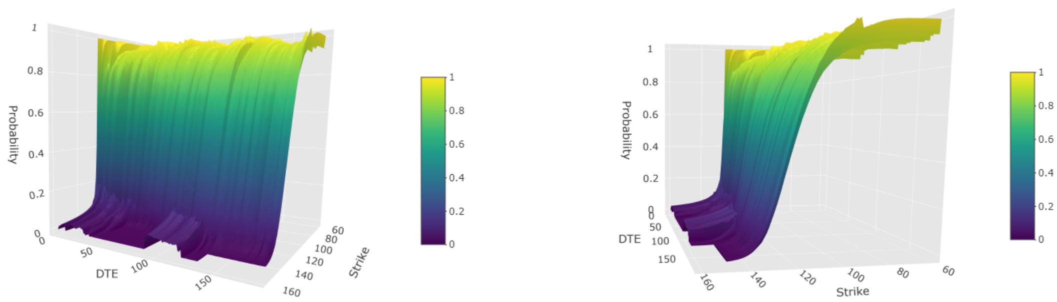


Figure 2: **Observed distribution of $N(d_2)$**

Note: Two views of a sample 3D plot of the observed distribution of $N(d_2)$ from BSM across 180 days. Dates with more vertical distributions are closer to the true distribution, which assigns less weight to for option contracts deep in and out of the money.

1.4 Results

For each unique options expiration date, we collect a sample of data points expanding 60/90/120/180 calendar days out from expiration. SPY prices are collected and used to create an empirical distribution step function for each test, which we then use to input strike prices from an options chain for each day. For each day prior to expiration, we bin the strike prices of the options chain according to the empirical distribution of the sample period. Hence, we form a forward ‘actual distribution’ of realized price outcomes for SPY as a base comparison to the BSM. We then calculate $N(d_2)$, for each day, using the described method in Section 3. Finally, we apply the non-parametric two sample K-S test, as described by Equation 6, which finds the largest point-wise difference between the two CDF curves. The p-value of the K-S test statistic, $D_k(F, G)$, is tested against a 5% level of significance. If the p-value is below 0.05, there is enough statistical evidence to reject the null hypothesis that both samples come from the same underlying population distribution[19]. We repeat a similar process for the A-D test, which differs from the K-S test in the location of power (comparing the tails of the distribution). In essence, we test the similarity between the implied distribution of BSM to the actual distribution using the prescribed goodness of fit measures from Section 2. Our emphasis is placed on how well the BSM forecasts the forward distribution of the underlying asset each day prior to expiration, and when we see the similarity occurring.

Figure 3 displays an extensive view of our research findings for option contracts that ex-

[19] Here, we assume both samples are log-normally distributed.

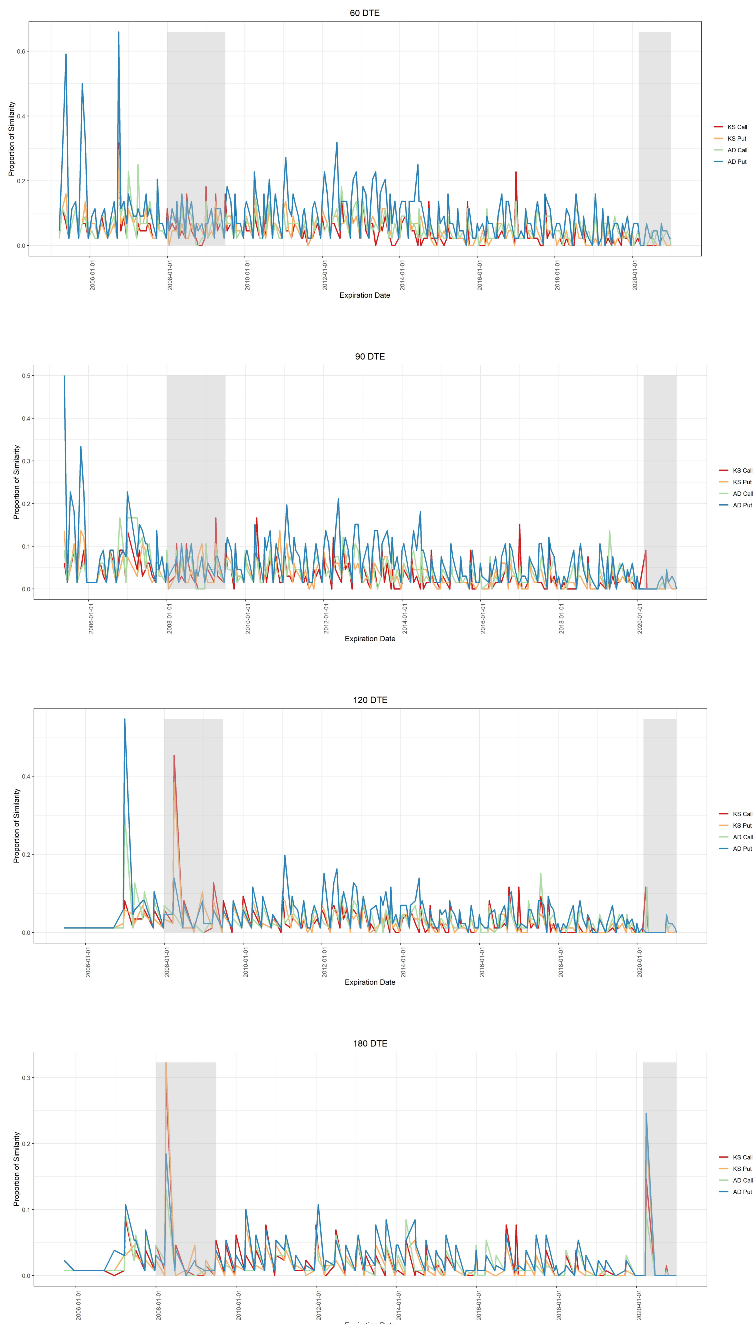


Figure 3: Summary Results for 60/90/120/180 DTE

Note: The proportion of similarity between the implied BSM distribution and the actual distribution from 60/90/120/180 days to expiration. The x-axis is all unique option expiration dates. The red(orange) line is the proportion of similarity for all call(put) option contracts under the K-S test. The green(blue) line is the proportion of similarity for all call(put) option contracts under the A-D test. Recessionary periods are highlighted in the gray scale of the plot, and does not include the full recessionary period of the COVID-19 period.

Table 3: Summary Results for Proportion of Similarity

(a) 60 DTE, $N = 246$

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
K-S Call	0	0.0227	0.0227	0.0433	0.0682	0.3182
K-S Put	0	0.0227	0.0455	0.0463	0.0682	0.2955
A-D Call	0	0.0227	0.0455	0.0580	0.0682	0.6591
A-D Put	0	0.0227	0.0682	0.0923	0.1364	0.6591

(b) 90 DTE, $N = 230$

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
K-S Call	0	0.0152	0.0152	0.0319	0.0455	0.1667
K-S Put	0	0.0152	0.0303	0.0327	0.0455	0.1364
A-D Call	0	0.0152	0.0303	0.0409	0.0606	0.1667
A-D Put	0	0.0152	0.0606	0.0642	0.0909	0.5000

(c) 120 DTE, $N = 191$

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
K-S Call	0	0	0.0116	0.0261	0.0349	0.4535
K-S Put	0	0	0.0116	0.0234	0.0349	0.3837
A-D Call	0	0.0116	0.0233	0.0300	0.0465	0.3023
A-D Put	0	0.0116	0.0349	0.0452	0.0698	0.5465

(d) 180 DTE, $N = 133$

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
K-S Call	0	0	0.0077	0.0189	0.0308	0.2923
K-S Put	0	0	0.0077	0.0166	0.0231	0.3231
A-D Call	0	0.0077	0.0077	0.0177	0.0231	0.1308
A-D Put	0	0.0077	0.0154	0.0269	0.0385	0.2462

Note: We summarize our results across each DTE, based on our sample of unique expiration dates. Here, N represents the total number of sampled expiration dates for which we conduct the GOF tests on. Using the proportion of similarity for each sampled expiration date, we provide descriptive statistics for the output of our GOF tests.

pire within 60/90/120/180 days. In each sub-figure, the x-axis represents all unique option expiration dates sorted for option contracts that expire within 60/90/120/180 days. The y-axis represents the proportion of similarity between the two distributions, that is, the total number of days we fail to reject the null hypothesis for the K-S or A-D test over the total number of DTE. We can then answer what proportion of the time, across 60/90/120/180 DTE, did the forecasted BSM probability distribution statistically predict the actual probability distribution. The red/yellow line in each sub-figure represent the proportion of similarity under the K-S statistic for call/put options with 60/90/120/180 DTE. The green/blue line in each figure represents the proportion of similarity under the A-D statistic for call/put options with 60/90/120/180 DTE. We can see from each of the sub figures within Figure 3 that for most expiration dates there is a very low proportion of similarity between the implied BSM distribution and the actual distribution.

Our results are summarized in Table 3, where we provide descriptive statistics for the proportion of similarity based on the sampled expiration dates with 60/90/120/180 DTE. We further extend our summary results in Table 4, where we take the average proportion of similarity for each option type, GOF measure, and across each expiration year in our sample. When looking at the difference in proportion of similarity between GOF tests as well as between option types, a few interesting results arise. For call options, the K-S test and A-D test show relatively equal proportions of similarity between the implied and actual distribution across unique expiration years. In contrast, the proportion of similarity for put options differs substantially between the two GOF tests; the A-D test fails to reject the null hypothesis more times than the K-S test. This is especially evident during the start of recessionary periods, and in shorter expiration dates of 60 and 90 days. For periods of 120 and 180 DTE, the K-S test suggests more similarity in the actual and implied distribution during the start of 2008. Notice, however, that the spike in proportional similarity occurs in the early years of the 2008 crisis. This is seen for expiration dates midway 2008, when financial markets began to decline and the VIX peaked. Downside exposure to price movements in the S&P 500 during this period would suggest that the implied probability distribution of put options are more similar to the underlying SPY price path than call options. Implied volatility on options typically increase at the onset of recessionary periods, resulting in an increase in option premiums and change in risk-adjusted probabilities. In the years following the '08 Financial Crisis, the proportion of similarity drops in comparison. This includes the most recent COVID-19 crisis, where the K-S and A-D test find very little similarity between the implied and actual distribution of SPY a part from a spike in the proportion of similarity for options with longer expiration dates of 180 days. One main difference in our results lies in the characterization of the two recessionary periods in our sample; the '08-'09 period stemmed from a decline in the market value of financial assets, whereas the most recent crisis was the result of a slowdown in economic activity during the shutdown of the global economy. From Tables 10 and 11, the implied volatility differential has seen a steady rise in the post '08 crisis period. This may suggest that there is a negative correlation in the steady rise in implied volatility differentials, and a lower forecasting ability in the BSM implied probability distribution.

Table 4: Mean Proportion of Similarity by Year

(a) 60 DTE

Year	N	KS Call	KS Put	AD Call	AD Put
2005	10	0.059	0.075	0.064	0.205
2006	14	0.081	0.075	0.115	0.131
2007	15	0.055	0.058	0.077	0.092
2008	16	0.055	0.041	0.061	0.077
2009	16	0.064	0.061	0.06	0.091
2010	16	0.057	0.063	0.064	0.107
2011	16	0.04	0.055	0.058	0.102
2012	16	0.067	0.064	0.077	0.136
2013	16	0.043	0.064	0.064	0.128
2014	16	0.045	0.053	0.06	0.107
2015	16	0.028	0.026	0.04	0.063
2016	16	0.04	0.027	0.047	0.067
2017	16	0.034	0.04	0.054	0.071
2018	15	0.015	0.02	0.041	0.059
2019	16	0.016	0.023	0.038	0.053
2020	16	0.004	0.011	0.018	0.034

(b) 90 DTE

Year	N	KS Call	KS Put	AD Call	AD Put
2005	8	0.047	0.074	0.049	0.189
2006	11	0.056	0.045	0.069	0.079
2007	13	0.048	0.047	0.075	0.08
2008	15	0.036	0.035	0.042	0.052
2009	14	0.05	0.047	0.044	0.061
2010	15	0.053	0.052	0.043	0.075
2011	16	0.029	0.039	0.046	0.075
2012	16	0.047	0.046	0.057	0.093
2013	15	0.025	0.036	0.04	0.085
2014	15	0.03	0.035	0.041	0.074
2015	15	0.021	0.017	0.029	0.045
2016	15	0.028	0.023	0.034	0.047
2017	16	0.023	0.023	0.037	0.054
2018	16	0.01	0.011	0.026	0.035
2019	15	0.015	0.015	0.029	0.038
2020	15	0.009	0.003	0.007	0.01

(c) 120 DTE

Year	N	KS Call	KS Put	AD Call	AD Put
2005	3	0.012	0.012	0.012	0.012
2006	5	0.026	0.023	0.07	0.128
2007	8	0.035	0.036	0.057	0.055
2008	8	0.076	0.077	0.025	0.049
2009	8	0.049	0.035	0.036	0.039
2010	8	0.039	0.033	0.029	0.048
2011	15	0.024	0.029	0.036	0.058
2012	15	0.042	0.034	0.043	0.078
2013	14	0.023	0.028	0.032	0.059
2014	15	0.026	0.027	0.033	0.055
2015	16	0.016	0.015	0.02	0.036
2016	15	0.033	0.019	0.026	0.037
2017	15	0.025	0.022	0.038	0.046
2018	16	0.006	0.007	0.019	0.027
2019	14	0.008	0.008	0.02	0.026
2020	16	0.012	0.001	0.012	0.007

(d) 180 DTE

Year	N	KS Call	KS Put	AD Call	AD Put
2005	2	0.008	0.008	0.008	0.015
2006	4	0.006	0.008	0.008	0.015
2007	8	0.032	0.026	0.038	0.045
2008	8	0.046	0.05	0.026	0.037
2009	8	0.022	0.012	0.016	0.02
2010	8	0.025	0.027	0.023	0.039
2011	8	0.026	0.019	0.027	0.028
2012	9	0.022	0.016	0.021	0.033
2013	9	0.018	0.023	0.017	0.04
2014	10	0.014	0.021	0.022	0.039
2015	10	0.015	0.01	0.015	0.023
2016	9	0.028	0.008	0.021	0.017
2017	9	0.01	0.009	0.009	0.024
2018	10	0.008	0.005	0.012	0.015
2019	9	0.003	0.002	0.006	0.009
2020	12	0.013	0.018	0.008	0.021

Note: Summary results for each year of our sample is provided in the table above, where N represents the number of options expiration dates in which we conduct our goodness of fit tests on. For each option expiration date, we calculate the proportion of trading days in which the actual and implied distributions are similar. Finally, we take an average of all expiration dates sampled in a given year.

These results are counter-intuitive to the notion that it is harder to achieve the critical value under the K-S statistic. Since the K-S test has less power in the tails of the distribution, we should see more failure to rejects under the K-S statistic; however, we see the opposite. A possible explanation for this may lie within the differing levels of implied volatility between put and call options. Implied volatility tends to be higher for put options than call options given that prices fall faster than they rise. The A-D test may be more sensitive to higher levels of volatility in the forecasted distribution than the K-S test, negatively affecting the rejection rate. Since the K-S test focuses on differences near the center of the distribution, much of the emphasis is placed on options closer to at-the-money. Since implied volatility is lower towards the center of the distribution, this may result in the underestimation of risk-adjusted probabilities. Overall, the results of Tables 3 and 4 suggest the BSM more accurately forecasts the probability distribution for put options than it does for call options. Differences between the two tests can also be attributed to varying increments of days to expiration, as the range of SPY prices in the sample used to construct the empirical distribution increases for longer dated options expiration dates. The average length of strike prices for an options chain, seen by Tables 1 and 15, are larger for maturities greater than 90 days to expiration. Hence, contracts deep out-the-money for longer dated maturities tend to be assigned a larger risk adjusted probability weight compared to shorter dated contracts, which feature a reduced number of strike prices.

After answering how often the BSM implied distribution equals the actual, the next question to ask is when BSM forecasts are most accurate. Figure 4 displays a slightly more granular view of our results for 120 DTE. Here, the x-axis is the amount of days to expiration starting from 120 out and the y-axis is the unique options expiration dates. As opposed to Figure 3, with Figure 4 we can see at which point in time the implied distribution from the BSM begins to accurately predict the risk-adjusted probabilities of the underlying price path. In addition, we can also observe the frequency of similarity between both distributions for each specific trading day. This is especially evident when looking across all options chains for contracts that have 60 DTE versus 180 DTE, which can be found in the **Appendix**. In Figure 4, a similar pattern emerges between option types and GOF tests; the probability distributions are most similar closer to expiration. In the early days of the option, beginning on trading day 120, there seems to be very little similarity in the distributions of risk adjusted probabilities. Specifically, about 5-10 trading days prior to expiration, the K-S and A-D test show distributional similarity between the implied and actual distributions. Closer to expiration, there is a shift in the assignment of risk adjusted probabilities away from deeper out-the-money contracts and into strike prices closer to the true underlying price path. That is, out the-money contracts have a significantly lower likelihood of being in-the-money within

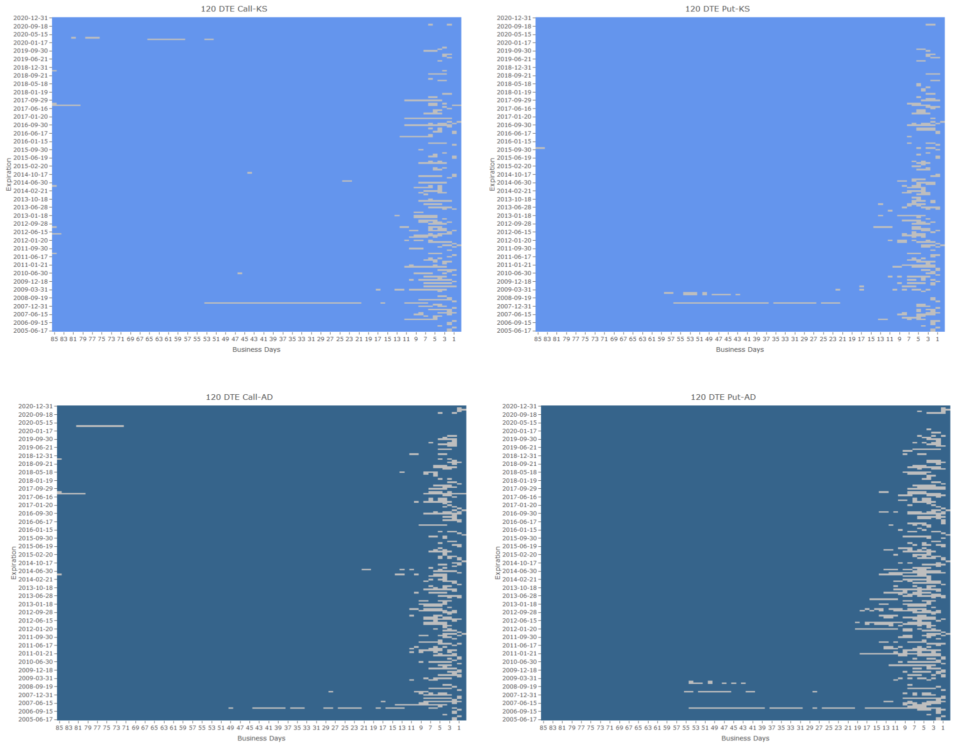


Figure 4: In Depth Summary Results for 120 DTE

Note: Implied BSM distribution and actual distribution tested under the K-S (top two sub-figures) and A-D (bottom two sub-figures) test for each unique expiration date across 120 days to expiration. The blue and red colors mean the null hypothesis was rejected on that day, meaning the two distributions do not come from the same underlying population distribution. The green and yellow colors mean we failed to reject the null hypothesis on a given day, meaning the two distributions are from the same underlying population distribution.

a short time frame prior to expiration.

Different from the K-S test, much of the distributional difference in the A-D test is determined in the tails. Far out from maturity, the BSM tends to overestimate the tails of the estimated CDF of $N(d_2)$. This is indicative of the BSM, again, having relatively weak predictive power in determining the price path of future SPY prices. Why might this be the case? The K-S test tends to overestimate the distributional similarity of the implied and actual probabilities, and the A-D test has more power than the standard K-S test comparison. Even so, the A-D test finds consistent results, compared to the K-S test, of distributional similarity between our two samples close to expiration. This result is backed by the visual representation of Figure[20] 4. The issue with the BSM is how probability measures are de-

[20]This also includes our visuals in the **Appendix** for Figures 8 , 9, and 10.

termined when factoring implied volatility[21]. With implied volatility being relatively higher for option contracts that are deep in- and out-of-the-money, the BSM overestimates the likelihood of the option being exercised above a given strike price. Kownatzki and Sabouni (2019) present the case of risk exposure to market downturns that limit the profitability of options trading strategies, such as in strangles. This is the case, as Kownatzki (2016) suggests, of the BSM exceedingly overestimating risk in periods of market normalcy. BSM assigns probabilities to a variety of strike prices in an options chain, even for deep in- and out-of-the-money contracts approaching expiration. Specifically, deep OTM contracts have a small likelihood of ever being exercised, but have significantly smaller option premiums compared to strike prices closer to ATM. In practice, traders may be net sellers of options due to the uncertainty underlying options farther out from expiration. Days prior to expiration, the need for OTM options to hedge large price swings is less needed. Hence, closer to maturity the BSM tends to predict risk-adjusted probabilities of SPY more accurately. A greater question arises about the implications of a short time horizon in the predictive power of the BSM: are traders better off being net holders of options closer to maturity? In days leading up to expiration, option premiums may be relatively expensive for strike prices with a greater probability of exercise ($N(d_2)$). Hence, profitability in such a strategy may be undermined by the cost of undertaking the options contract. Implied volatility may also be relatively high in shorted dated options, especially for ATM and deep ITM options illustrated by Figure 12. Undertaking a long position in a contract may be difficult and risky to undertake in the short-run, even in light of distributional similarity of market expectations implied by the BSM and underlying price.

1.5 Conclusion

The BSM framework has been a key tool in modeling derivatives, and the underlying characteristics of a security. We used the BSM to model the predictability of options in determining the price path of SPY by estimating $N(d_2)$. Our results suggest the market expectations as backed out by the BSM falls short in forecasting the price path of SPY, which we determine by conducting a two sample K-S and A-D GOF test. In shorter time frames particularly prior to expiration, we begin to see a similarity in the two sample distributions. Much of the probability density in the distribution of $N(d_2)$ seems to be overestimated by the BSM, suggesting the distribution is fat tailed for options deep in and out the-money. This issue

[21] Issues such as non-stationarity in volatility may be one main issue with the BSM overestimating implied probability distributions.

of non-normality in financial time-series data is common and is evident in our analysis of the BSM when assessing risk-adjusted probabilities. Nonetheless, we examine a unique phenomenon of changing risk-adjusted probability measures. Through time, the shape of the CDF of $N(d_2)$ on an options chain changes. This change can be characterized by traders assessing less and less probability for tail events to occur the closer we get to expiration. Hence, strike prices within a narrow band closest to the underlying spot price are much more likely to occur closer to expiration.

The K-S and A-D tests provide a non-parametric approach to modeling the forecasted likelihood of options expiring in the-money. We find that the resulting distributions differ considerably, specifically in time horizons greater than 40-60 days from expiration. Much of this is attributed to the volatility smirk on options, where deep in- and out-of-the-money options have the highest measures of implied volatility. This difference diminishes as time passes, where a reduction in the time value component of the option results in a more accurate measure of option prices and risk. We find this to be the case in our analysis, where the probability distribution from BSM reflects the underlying price path of SPY better as an option approaches maturity.

We do not find the BSM to be a good forecasting tool for longer time horizons: when options are far from expiration. Instead, greater research is needed on how risk-adjusted probabilities from $N(d_2)$ are affected by factors such as implied volatility. We see from a Monte Carlo simulation of assuming constant volatility that the probability of tail events are overestimated from the BSM. This is because implied volatility is allowed to vary across an options chain, assigning higher risk to certain strike prices. Longer dated options seem to not reflect market expectations on the future underlying price path, based on the likelihood of exercise determined by $N(d_2)$. This reflects the inability of traders to accurately predict future price movements, and indeed overall market trends.

2 Chapter 2: Forecast Error: A Cause of Government Intervention or Market Conditions?

Coauthored with Hisam Sabouni

2.1 Introduction

Often, economists are tasked with creating models to forecast the state of the economy. Forecasting in macroeconomics is inherently challenging, as often economists do not have a clear sense of the current state of the world to make a solid judgement about the future state of the world. For instance, most macroeconomic data series arrive at slow intervals and are often subject to substantial revisions. There is improvement on the speed of macroeconomic data through the advent of high frequency indicators that provide near real time insights into the health of the economy Sabouni (2018), Cavallo and Rigobon (2016), Choi and Varian (2009)).

The Survey of Professional Forecasters, is a quarterly survey conducted by the American Statistical Association (ASA) and the National Bureau of Economic Research (NBER) from 1968:Q4 until taken over by the Federal Reserve Bank of Philadelphia in 1990:Q2 [Croushore (1993)]. The Survey of Professional Forecasters is the oldest quarterly survey of macroeconomic forecasters in the U.S. The SPF is sent to various professional economists which gains insight on their views about the economy over the next few years, and is considered to be one of the best aggregations of economic forecasts [Zarnowitz and Braun (1993)]. The survey mostly asks for point forecasts, for a range of variables (such as U.S business indicators, implied forecasts and macroeconomic indicators) and time horizons, but also requests a density forecasts [21]. Diebold et al. (1997) assessed the adequacy of SPF forecasts, specifically the density forecasts for output and inflation, and found SPF forecasts suffered from overestimation[22] as well as serial correlation[23]. In contrast to Diebold et al. (1997), we are aware that the SPF forecasts are somewhat inaccurate and analyze which type of inputs are largely responsible for increases in SPF forecast error.

The purpose of our paper is to study the characteristics of error in economic forecasts over time and analyze whether the forecast error from SPF is mainly driven by two factors:

[21] Density forecasts are requested for output and inflation. Forecasts were requested for nominal inflation, until they switched to asking survey respondents to provide forecasts for real output in the early 1980s.

[22] For the probability of large negative inflation shocks.

[23] Present in inflation surprises.

macroeconomic conditions or government intervention via monetary and fiscal policy. We focus on explaining the variation in errors of SPF forecasts across three financial securities by isolating the effects of changes in fiscal and monetary policy as well as changes in various macroeconomic indicators. While very little is known about the construction of the point estimates and density forecasts reported by survey participants Diebold et al. (1997), we can analyze whether the driving force of SPF forecast error for various securities stems from changes in government policy or changes in macroeconomic indicators (or macroeconomic conditions such as changes in). Each survey participant may use a different forecasting model, but our analysis will still shed light on which group of variables is most likely throwing off their estimates. We analyze the average output of survey participants and concentrate on groups of inputs. Using principal component analysis, we reduce the dimensions of these macroeconomic variables and interpret the first two principal components of the macro PCA as the overall effect macroeconomic conditions has on forecast error. Using a linear regression, we test whether monetary policy, fiscal policy, or the combination of the two, are responsible for increases in the SPF's forecast error of each security. Our goal is to isolate the effects on SPF forecasts of changes in fiscal and monetary policy, as well as changes in other macroeconomic indicators. We hypothesize government intervention greatly impacts the expectations of professional forecasters.

We find that increases in monetary policy largely affects the short-term security used in our analysis, with limited economic impact to the forecast error of the two longer-dated securities. Fiscal policy has a statistically significant effect, but relatively insignificant economic effect, on the forecast error of the included longer-dated securities. Increases in macroeconomic conditions consistently has a statistically significant, and rather significant economic effect, on the forecast error of all three securities in our analysis. While many papers focus on the ex-post evaluation of ex-ante forecasts, we contribute to existing literature by examining which groups of inputs may largely be contributing to SPF forecast error. Forecasters may be able to use this information to adjust their expectations surrounding future changes in government policy or macroeconomic conditions. We also pull from existing machine learning literature in utilizing sub-models to explain baseline model errors.

The remainder of the paper is organized as follows: Section 2 provides a brief literature review regarding the effect of government policy within the U.S. financial market. Section 3 summarizes the data. In Section 4, we describe the methods used to reduce the dimensionality of our macroeconomic data as well as introduce our formal statistical model. Section 5 presents and discusses our results. The last section concludes.

2.2 Government Policy's Impact in Financial Markets

When investors are uncertain about the future, they will generally feel less safe tying up their capital in longer term securities such as 10-year Treasury bonds. Analyzing a bond's yield across various maturities to form what is known as a bond's yield curve provides a good measure of future economic growth expectations. The relationship between long and short-term yields (ex. 10-year and 3-month) helps measure overall market uncertainty Tsatsaronis and Smets (1997). As a result, many investors analyze the steepness of the yield curve as a proxy for other investor preferences. In fact, by imposing no-arbitrage conditions, an individual can use yield curve data to back out what market expectations are on future interest rates. Therefore, accurate forecasting of the implied yield is crucially important to generate precise market predictions. However, through no-arbitrage restrictions, government intervention, such as monetary policy, has the ability to affect the entire term structure of interest rates since the actions of the Federal Reserve at the short end of the yield curve influence the dynamics of the long end of the yield curve Ang et al. (2011).

While the yield curve is directly tied to the actions of the federal governments actions (fiscal policy), it is monetary policy that influences the slope of the yield curve (see Bomfim (2003), Mumtaz and Surico (2008), Brand et al. (2006)). When there is an increase in monetary policy via instruments such as the effective federal funds rate (EFFR) by the Federal Reserve, this leads to a rise in short-term interest rates to reduce inflationary pressures. The economy will experience a slowdown (rGDP decreases) and the yield curve will flatten. The actions of the Federal Reserve affect short-term interest rates differently compared to long-term interest rates (see Wood (1964) Haldane and Read (2000), Kuttner (2000)). The effect of monetary tightening as previously described is the effect on short-term interest rates. For long-term interest rates, the tightening effects the market's future expectations about inflation which affects the rate investors are willing to accept on longer term debts. In contrast, an increase in fiscal policy would increase expectations of inflation in the long-term. Investors will demand greater yield at the long end of the curve to compensate for the diminishing value of money. When analyzing the effect of both monetary and fiscal policy on the nominal Treasury yield curve, Evans and Marshall (2007) found monetary policy was an important transmission pathway for both technology shocks and shocks to preferences for current consumption on interest rate variability, while limited evidence supported the importance of fiscal policy.

In addition to monetary and fiscal policy, macroeconomic indicators also can have a large impact on forecasting the implied yield as the dynamics of yields are tied to the dynamics of macroeconomic variables. Macroeconomic variables and yields are both characterized by a high degree of co-movement since a large part of their dynamics are driven by common forces**?**. Yields that have a maturity of less than one year move closely with the policy instrument used by central banks; which responds to changes in inflation, economic activity, or other economic conditions Taylor (993a). The average level of the yield curve is correlated with the inflation rate and the spread between long and short rates with temporary business cycles conditions Diebold et al. (2004). The short-term interest rate is set by the central bank according to its macroeconomic stabilization goals. Long-term yields are largely determined by expectations of future short-term interest rates, which in turn depend on expectations of the macroeconomic variables Diebold and Rudebusch (2013). Additionally, macroeconomic variables have a strong effect on future movements of the yield curve Diebold et al. (2004). Given this relationship, the inclusion of macroeconomic variables has improved the accuracy of forecasting future interest rates Ang and Piazzesi (2003). Duffee (2013) shows macro-finance models can be used to improve forecasting security yields that are out-of-sample by reducing the problem of over-fitting. In this paper, we analyze for various securities whether it is monetary policy, fiscal policy and/or macroeconomic indicators that largely effects SPF forecast error.

2.3 Data

This study focuses on three main securities; the three-month Treasury bill, the ten-year Treasury note and Moody's AAA corporate bond. We obtain quarterly yields for each security from Bloomberg. The sample period is from 1980 to 2018, except for Moody's AAA corporate bond, an index of AAA rated corporate bonds[24], which had data available from 1983 to 2018. Quarterly forecast data for these three securities was obtained from the Survey of Professional Forecasters (SPF) by the Federal Reserve Bank of Philadelphia. Forecasters provided quarterly average projections for one to six quarters ahead[25]. For each security, we construct the average forecast at each quarterly projection as the simple average of forecasters. Using a time trend variable, we are then able to stack each of our six models on top of each other for our analysis; essentially averaging across six epsilons[26]. By doing so, we increase the total amount of forecast data available, which is needed at this data frequency.

[24]This index tracks the average yield for companies with AAA ratings by Moody

[25]Except for Moody's AAA corporate bond, which SPF did not have yield projections for one quarter ahead.

[26]Except for Moody's AAA corporate bond, which will be averaging across five errors, or epsilons, in our analysis.

Table 5: **Summary Statistics**

Panel A: Forecast Error	Mean	SD	1st Q	3rd Q
3-Month T-Bill FE	-0.359	1.348	-0.922	0.430
Moody's AAA C-Bond FE	-0.436	0.852	-0.855	0.011
10-Year T-Bond FE	-0.380	0.771	-0.904	0.139

Panel B: Variables that Largely Affect Forecast Error	Mean	SD	1st Q	3rd Q
Effective Federal Funds Rate	4.587	3.944	1.038	6.618
US Govt Receipts	12.736	3.725	9.597	15.667
US Govt Outlays	13.474	3.672	11.026	15.075
Real Gross Domestic Product	2.786	2.724	1.500	4.125
Personal Consumption Expenditures	2.701	1.697	1.675	3.450
Federal Debt	65.429	21.192	52.553	78.085
Unemployment Rate	6.257	1.669	5.000	7.300

Note: Panel A displays summary statistics on the SPF forecast error from 1980 Q4 to 2018 Q3 for three-month Treasury bill and ten-year Treasury note and from 1983 Q1 to 2018 Q3 for Moody's AAA corporate bond. Forecasts are the average of all unique forecasts for a given quarter. Forecast error is actual yield minus the average forecasted yield across all 6 projections for a given quarter. Panel B displays summary statistics on the variables that largely affect forecast error from 1980-2018. Each variable listed in Panel A and Panel B of 7 is expressed as a rate.

For a timeline and more detailed explanation of forecast projections, see the **Appendix**. To ensure data quality, we require there to be a minimum of 2 unique forecasters to average across for a given quarter, counted by unique analyst using SPF (Id). After analyzing the data, we found there is at least 9 or more unique forecasts made each quarter between 1980 Q1 to 2018 Q3. This adds consistency to our average quarterly forecasts.

We use the effective federal funds rate as measure for monetary policy from the Federal Reserve of Economic Data (FRED). As a measure for fiscal policy, we use receipts and outlays. Monthly receipt and outlay data were collected from the Monthly Treasury Statement provided by the Fiscal Data.Treasury.gov. As a measure of macroeconomic conditions, we utilize several macroeconomic variables. Quarterly real gross domestic product (rGDP), personal consumption expenditures (PCE), total federal debt and the unemployment rate for the U.S was collected from FRED. We convert receipts, outlays and total federal debt as a percentage of real GDP.

Table 7 shows descriptive statistics of our dependent variable, forecast error, for each of the three securities. We define forecast error in this paper as the actual yield minus the average SPF forecasted yield across all 6 projections for a given quarter. Forecast error statistics for each projection separately can be found in the **Appendix**. We additionally show in Table 7 descriptive statistics on variables that largely affect forecast error of the yield curve: government policy instruments and macroeconomic indicators. Panel A reveals SPF forecast error for all three securities is negative, meaning on average, forecasts are overestimated. Average forecast error is largest for Moody's AAA Corporate Bond, while the two treasury securities, the three-month treasury bill and ten-year treasury note, have relatively equal levels. However, between all three securities, forecast error was spread out over the largest range of values for the three-month treasury bill with a standard deviation of 1.348 percentage points. This could be a result of yields varying so much in the short-run in comparison to the long-run. Panel B summarizes key variables that are responsible for changes in the forecast error of a securities' yield or yield curve. The first three variables under Panel B are government policy instruments and the remaining variables are macroeconomic indicator variables. Each of these variables is expressed as a rate or has been converted into a rate (as a percentage of GDP).

Figure 5 displays average SPF forecast error across time for each of the three securities. The black line represents the average error for forecasts made 1 quarter into the future to 6 quarters into the future. The red dotted lines represent the 1st and 3rd quartiles; meaning, 50% of the SPF forecasts are within these two red dotted lines. From Figure 5 we can see the black line for each of the three securities increases as we forecast further into the future. This makes sense because there is more uncertainty the further we go out in time. Average SPF forecast error increases the sharpest across the 6 forecast horizons for the three month Treasury bill. We can also notice in the left graph for the three month Treasury bill as forecasts are made further into the future, the red dotted lines separating further from black line. This means the breadth of the forecast error distribution is wider (forecasts are more spread out, not close around the mean) for SPF forecasts made 4-6 qtrs. into future compared to the other two securities.

2.4 Methods

2.4.1 Dimension Reduction with PCA

Since we are using quarterly data, we have a limit on degrees of freedom. To address this issue, we apply an orthogonal dimension reduction on the four macroeconomic variables which we do not need independent effects for. By doing so, we will effectively reduce dimensionality or complexity of our models while sacrificing only a small portion of the total variation

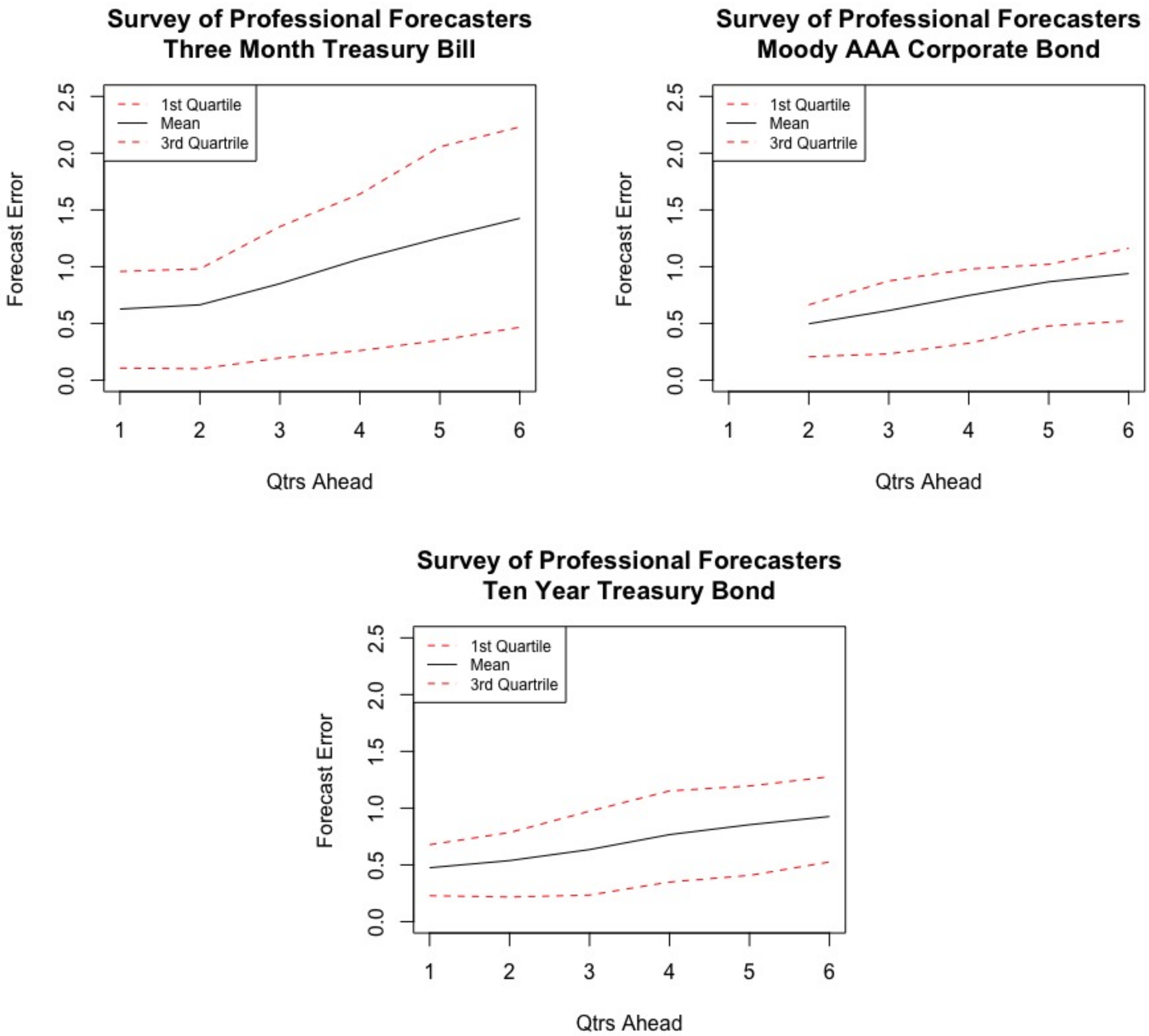


Figure 5: Average SPF Forecast Error Across Each Quarterly Projection Horizon
Note:: The left figure displays SPF forecast error across quarterly projections for the three-month Treasury bill. The middle figure displays SPF forecast error across quarterly projections for Moody's AAA corporate bond. Lastly, the figure on the right displays SPF forecast error across quarterly projections for ten-year Treasury note. The solid black line represents the average SPF forecast error across quarter projections. The two outer red dotted lines represent the 1st and 3rd quantiles of the average forecast error made at each quarterly projection.

from all four macroeconomic variables. In addition to solving the issue of limited degrees of freedom, PCA is also a technique for analyzing multiple regression data that suffer from multicollinearity, which macroeconomic data generally does, since each principal component (PC) vector is a linear combination of all the variables and are orthogonal to one another **?**.

After determining the appropriate amount of PCs to include, a trade off between explained variation and dimension reduction, we then run a principal components regression (PCR). Using a PCR, we analyze the effect of the EFFR, receipts and outlays (our measures of government policy) jointly with the collapsed vector of macroeconomic variables on SPF forecast error by means of a multivariate unobserved components time series model that is represented as a linear Gaussian state space framework.

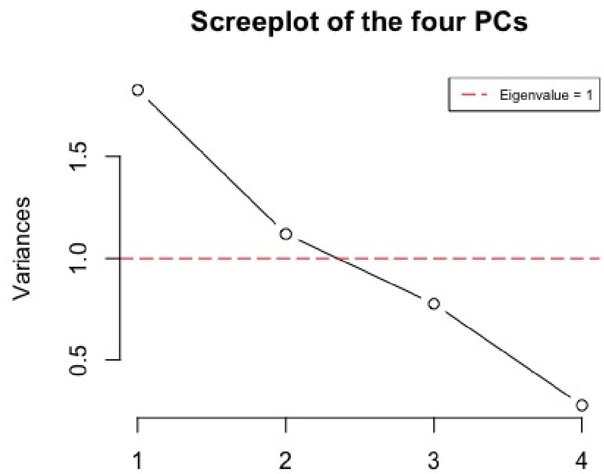
To determine the appropriate amount of PCs to include in our PCRs, we utilize a scree plot which plots the eigenvalues of factors or PCs in our analysis. The red horizontal dotted line in the left plot of Figure 2 represents an eigenvalue equal to 1. An eigenvalue < 1 would mean that the component actually explains less than controlling for a single explanatory variable; therefore, we would like to discard those. Since only the first two PCs have an eigenvalue equal to or greater than one, we will discard the third and fourth components. The trade off to reducing dimension as previously mentioned is a "loss" in the total variance from using all four components or all four explanatory variables. The plot on the right in Figure 2 shows the cumulative amount of variation explained from including an additional PC. The vertical blue dashed line represents at which PC is our cutoff and the horizontal line represents the corresponding cumulative amount of explained variance for that number of PCs. When we discard two components, we retain around 73% of the variance. Using two PCs, we can effectively reduce dimensionality from four to two while only "loosing" about 17% of variance.

2.4.2 Formal Testing

We utilize three PCRs to analyze the effect macroeconomic conditions and government policy have on SPF forecast error, one for each security. We do not include all three securities into a panel regression to get the average effect across all three securities for a couple reasons, such as differences in maturity lengths and liquidity levels. For those reasons, the following PCR specification is applied to the each of the three securities:

$$\begin{aligned} ForecastError_t = \beta_0 + \beta_1 EFFR_t + \beta_2 Receipt_t + \beta_3 Outlay_t + \beta_4 PC1_t \\ + \beta_5 PC2_t + \beta_6 Time_t + \beta_7 Actual_{t-1} + \epsilon_t \end{aligned} \tag{10}$$

In equation (1), $ForecastError_t$ is the actual yield minus the SPF forecasted yield for the given security; where the forecasted yield for a given year/qtr. is the average of all unique forecasts for that year/qtr. $PC1_t$ and $PC2_t$ are the first two PCs from the macroeconomic



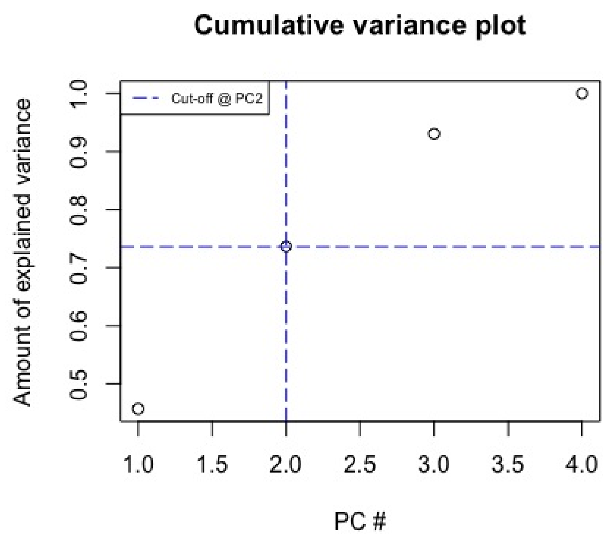


Figure 6: Scree and Cumulative Variation Plot of Macroeconomic Variables
Note: The top figure shows a scree plot used to determine the number of factors to retain in our PCR. The x-axis represents each PC. The y-axis is the variance. The red dotted line indicates an eigenvalue equal to one. We discard components whose eigenvalue is equal to less than one. The bottom figure shows the cumulative variance as we include an additional component. The x-axis represents each PC. The y-axis is the amount of variance the principal component vector accounts for. The blue-dotted line is the cut-off of PCs we include in our analysis. Using only two PCs, we can explain roughly 73% of the total variance while reducing dimensionality.

PCA we described in Section 3.1. $Time_t$ is a discrete variable from 1 to 6, controlling for length of the forecast projection. By just putting time in the regression, we are controlling for the linear relationship between forecast time and forecast error. $Actual_{t-1}$ is previous quarters actual yield for the given security.

2.5 Results

Table 6 displays the results for the average effect that government monetary and fiscal policy (via EFFR, receipts and outlays) and macroeconomic conditions (via macroeconomic indicators) have on SPF forecast error for our three main securities. Using the specification outlined in equation 10, we are able to see the 1:1 change, (immediate effect) a change in monetary policy, fiscal policy or macroeconomic conditions has on SPF forecast error for each of the three securities. That is, the change in government policy or macroeconomic indicators would be what is missing in the forecasts and hence driving the increase in forecast error. Since forecasts are made 1-6 quarters ahead of the realized yields, the forecasts may not be wrong. Instead, the error may be coming from policy changes made during these 1-6 quarters after the forecast was made. From table 6 we can see a 1% increase in the EFFR (monetary policy) leads to on average, a 1.028 percentage point increase in SPF forecast error for the three-month treasury bill. In contrast to this short-term security, a 1% increase in the EFFR leads to on average a 0.13 percentage point decrease in SPF forecast error for Moody's AAA bond and a 0.094 percentage point decrease for the ten-year treasury note. Interestingly, increases in monetary policy had statistically significant effects to SPF forecast error of Treasury type securities, while SPF forecast error of corporate bonds suffered minimally.

Turning to the effect of fiscal policy, we can examine two average effects: what happens to SPF forecast error with increases in government spending and what happens to SPF forecast error with increases in government collections (i.e. taxes). A 1% increase in receipts (government collections) leads to on average, a 0.003 percentage point increase in SPF forecast of the three-month Treasury bill, a 0.058 percentage point increase for Moody's AAA bond and a 0.021 percentage point increase for the ten-year Treasury note, respectively. The estimated coefficient for the three-month Treasury bill is small and statistically insignificant. In contrast to collections, a 1% increase in outlays (government spending) leads to on average, a 0.054 percentage point increase in SPF forecast error of the three-month Treasury bill, a 0.029 percentage point increase for Moody's AAA bond and a 0.019 percentage point increase for the ten-year Treasury note, respectively. The estimated coefficient for the ten-year Treasury note is small and statistically insignificant. These results suggest changes in government collections largely effect the forecast error of longer term securities while government

Table 6: Regression Output

	Dependent variable: Forecast Error		
	Three Month Treasury Bill	Moody's AAA Corporate Bond	Ten Year Treasury Bond
EFFR	1.028***	−0.013	−0.094***
	(0.058)	(0.028)	(0.030)
Receipt	0.003	0.058***	0.021**
	(0.013)	(0.012)	(0.011)
Outlay	0.054***	0.029**	0.019
	(0.015)	(0.014)	(0.012)
Time	−0.142***	−0.129***	−0.125***
	(0.019)	(0.021)	(0.016)
PC1	0.684***	0.340***	0.439***
	(0.062)	(0.060)	(0.075)
PC2	−0.195***	0.167***	−0.167***
	(0.042)	(0.042)	(0.037)
Actual_{t-1}	−0.812***	0.279***	0.296***
	(0.060)	(0.046)	(0.049)
Constant	−1.794***	−2.983***	−1.873***
	(0.315)	(0.466)	(0.361)
Observations	873	695	621
R^2	0.474	0.151	0.214
Adjusted R^2	0.470	0.143	0.205
Residual Std. Error	0.981 (df = 865)	0.789 (df = 687)	0.688 (df = 613)
F Statistic	111.386*** (df = 7; 865)	17.486*** (df = 7; 687)	23.824*** (df = 7; 613)

Note: This table summarizes the average effects of U.S. monetary policy via EFFR, U.S. fiscal policy via government receipts and outlays and overall market conditions via the 1st and 2nd PCs of the PCA. These results are presented for the three-month Treasury bill, Moody's AAA corporate bond and the ten-year Treasury bond.
*p<0.1; **p<0.05; ***p<0.01

spending largely impacts the forecast error of short-term securities in our analysis.

Increases in overall macroeconomic conditions also negatively impacts SPF forecast error. PC1 is the 1st principal component we included in our PCR, which is comprised of a linear combination of 4 macroeconomic variables and accounts for most of the total variation between these variables (as descried in Section 4.1). A 1% increase in overall macroeconomic conditions (PC1) leads to on average, a 0.684 percentage point increase in SPF forecast of the three-month Treasury bill, a 0.340 percentage point increase for Moody's AAA bond and a 0.439 percentage point increase for the ten-year Treasury note, respectively. These results suggest shocks to macroeconomic indicators are hard to forecast; these changes may be more unexpected and thus difficult to form accurate expectations. While the average effect of monetary policy via EFFR on the three-month Treasury bill is largest across all securities and between all variable types, increases in macroeconomic conditions affect all security types to a relatively high degree. The second PC, which captures the second largest unique[27] variance, on average decreases SPF forecast error for the two Treasury securities. The models for each of the three securities, especially for the three-month Treasury bill, have relatively high R^2, which supports the notion that future changes to government policy and macroeconomic indicators largely can explain the variation in SPF forecast error.

2.6 Conclusion

The SPF is comprised of an expert panel of forecasters and has been around for several years providing highly valued forecasts. We directly test the variation in errors of SPF forecasts for three main securities and estimate whether future changes in government policy or macroeconomic conditions is the driving force of these errors by impacting the expectations of professional forecasters. Distinguishing the role government policy plays in SPF forecast error provides insight into how powerful a shock to monetary or fiscal policy can be within the US financial market as well as for macroeconomic shocks.

Since we have quarterly data, we utilize machine learning dimension reduction techniques in order to preserve degrees of freedom and enhance the consistency of our estimates. We reduce dimensionality by discarding two PCs while still retaining 73% of our macroeconomic explanatory variables. The remaining two PCs are a linear combination of all four macroeconomic variables, which can be interpreted as a vector of overall macroeconomic conditions. We increase the number of forecast observations in our study by essentially stacking six models on top of one another. In other words, we effectively have six ϵ's; one for each forecast horizon. Average SPF forecast error increases the further the forecasts extend, consistent with the notion there is more uncertainty as we go further into the future. Average SPF forecasts tend to be overestimated for each of the securities.

We find SPF forecasts are highly sensitive to changes to both input groups that we include in our study, but forecasts were more consistently sensitive across security types to changes in macroeconomic conditions. In regard to government policy, increases in monetary policy via the EFFR have especially large impacts on SPF forecast error. Investors may perceive forecasts for short-term securities as less accurate when the expectations for future changes in monetary policy are high. Overall, our results suggest changes in government policy affect the short-term security in our analysis to a large magnitude, but SPF forecasts are largely sensitive more consistently to changes in macroeconomic conditions across the three securities.

[27] The first eigenvector is orthogonal to the second eigenvector as to maximize the variance accounted for.

3 Chapter 3: Seasonal Decomposition of Abnormal Market Returns

3.1 Introduction

The role of capital markets is to signal information about the relative value of goods and services. For example, a firm announces their end-of-quarter earnings and investors buy, sell, or do nothing based on that information. If there is novel favorable information within the announcement, investors may start buying this firm's stock and if there is novel unfavorable information, investors may instead sell. When investors react irrationally to this information, the signal from capital markets becomes distorted. The Efficient Market Hypothesis (EMH) proposes prices fully reflect all available information and that new information is primarily responsible for stock price movements Fama (1969); however, research in the 1980's, which has come to form what is now known as behavioral finance, found several financial market anomalies that violate the EMH. Some popular examples of these financial market anomalies are the January Effect, which is a pattern that shows higher returns tend to be earned in the first month of the year that remains unexplained after controlling for tax induced transactions. Or the Weekend Effect, which identified stock returns on Mondays are often significantly lower than those of the immediately preceding Friday, possibly due to the tendency of firms releasing bad news on a Friday after the markets close, which then depresses stock prices on Monday Thaler (1987). From these examples it becomes quite clear that Fama (1969) was correct, "Like any other extreme null hypothesis, I do not expect it [the EMH] to be literally true". Markets are neither fully efficient nor fully inefficient at pricing information, which can lead to investment opportunities for investors.

The question then becomes, when do markets price information correctly and when do they not? Is it the type of information that matters, or the timing of when that information is released that causes market inefficiency? Generally, information that is readily available and easy to interpret is more likely to be incorporated into the market efficiently by investors. Since markets are neither fully efficient nor fully inefficient, we can expect individual securities to have a mixture of correctly and incorrectly priced information. For example, Roll (1984) analyzed price fluctuations in the market for orange juice and found news about weather conditions, a large determinant of the supply of oranges, can explain only a small fraction of the variation in returns. weather related to orange juice production is easy to measure, reported accurately and consistently by the National weather Service of the

Department of Commerce. So, while weather-related news may be incorporated efficiently, other types of information are not.

Challenging the view that price movements are wholly attributable to the arrival of new information, this paper analyzes anomalies within the US financial market in the context of earnings announcements. Frazzini (2007) found stock prices rise around earnings announcement dates and that said price increase is strongly related to the fact that volume surges around the announcement date. Stocks with high volume around earnings announcements had both high premiums and a spike in buying by individual investors, suggesting prices rise around announcement dates due to buying pressure from investors. Savor et al. (2016) offers a risk-based explanation for the earnings announcement premium. Savor et al. (2016) shows investors use earnings announcements to adjust their performance expectations of non-announcing firms. In result of this, the co-variance between firm-specific and market cash-flow news spikes around announcements. This means although a firm's market beta (risk) may increase on the earnings announcement date, the increase in its expected return will be larger than can be explained by its risk alone. Savor et al. (2016) expects a positive announcement return even if the difference between news and expectations of earnings by investors is zero.

These high premiums and spikes in investor trading behavior following an earnings announcement leads to abnormal returns. Earnings announcements fit the criteria of being readily available and contain easy to interpret information, meaning they should be more likely to be efficiently be incorporated into the market. However, with earnings announcements, the difference between the realization of the announcement and investor's expectations comes as an unexpected shock resulting in the immediate buying/selling of securities and results in abnormal returns (see Ball and Brown (2014), Chambers and Penman (1984)). Engelberg et al. (2018) found abnormal returns are six times higher on earnings announcement days which can be associated with investor's biased expectations. Linnainmaa and Zhang (2019) documents abnormal returns are accumulated at different rates throughout the announcement period. Stocks earn significantly negative abnormal returns before earnings announcements and positive abnormal returns after the announcement. Interestingly, Linnainmaa and Zhang (2019) found these abnormal return patterns are unrelated to the earnings announcement premium.

Studies surrounding the effect of earnings announcements on stock returns are even greater importance as the information content within earnings announcements has been increasing for quite some time. Landsman and Maydew (2002) et al. found an increase in the informativeness of quarterly earnings announcements from 1969-1999. Francis et al. (2002) et al. believes this increase is due to expanded concurrent disclosures in firms' earnings announcements, specifically, the inclusion of detailed income statements. More recently, Beaver et al. (2017) et al. found there is a rising increase in information content of quarterly earnings announcements. They found one-day announcement windows exhibited roughly twice the price response observed for three-day windows; measured as the absolute magnitude of stock price revision at earnings announcements relative to price revision at other times from 1999 to 2012. While earnings announcements are not responsible for the day-to-day fluctuations in returns, they do play a large role when a firm experiences an extremely large stock price movement. I hypothesize it is not random occurrence firms experience this extreme variation, rather it is largely driven by investor's behavior to earnings surprises and that behavior is asymmetrically and seasonally dependent. In order to demonstrate this, I analyze the role earnings surprises have on abnormal returns.

Abnormal returns are defined in this paper as the actual return minus the expected return. In traditional event studies, expected returns are generally calculated by a market model, the capital asset pricing model (CAPM), the Fama French three factor model or any of the Fama French model extensions. These traditional event study models assume volatility or systematic risk of a security, compared to the market, is homoskedastic throughout the announcement period; however, the price reaction that occurs following an earnings surprise results in an increase to the variability and co-variability of securities returns Ball and Kothari (1991). Not allowing risk to vary decreases the ability to distinguish between increased expected returns and actual abnormal returns Ball and Kothari (1991). Given that stock returns generally exhibit time-varying volatility, a model which adjusts for risk may provide more efficient estimates in the calculation of abnormal returns. Ball and Kothari (1991) found when adjusting for risk increases at earnings announcements, firm's still accumulated abnormal returns. Using a conditional heteroskedastic model, such as a general auto-regressive conditional heteroskedastic (GARCH) model, relaxes the restriction that systematic risk of returns is constant during the announcement period. More specifically, a GARCH-in-mean (GARCH-M) model adds a heteroskedasticity term into the mean equation, which allows risk to vary across time. I use a CAPM-GARCH(1,1)-M model which I believe can more accurately distinguish between an increase in expected returns and true abnormal returns given that a GARCH model has the ability to capture the volatility clustering that prices exhibit Bollerslev et al. (1986) Engle (1982) Ball and Kothari (1991).

Using a fixed effects linear regression, I test the role earnings surprises play in generating abnormal returns by estimating whether investors' reactions to earnings surprises are asymmetric and/or seasonal. While we expect investors to react to announcements, we do not expect their reaction to be dependent on the month of the announcement. In fact, any seasonal patterns that persistently exist in financial markets should be arbitraged away by financial incentives of market participants. Although limits to arbitrage could result in changes to returns, it is not responsible for the behavior of investors to these announcements. If limits to arbitrage was responsible for these changes, investors would simply exploit this seasonality. I document that there does exist heterogeneous reactions from investors dependent on the earnings announcement sign and month. An earnings beat on average causes a 0.0017 percentage point (340% [28]) increase in daily abnormal returns. In contrast, an earnings miss on average causes a 0.0006 percentage point (120% [29]) decrease in daily abnormal returns. These results suggest negative earnings surprises on average have a limited impact on returns. By seasonally decomposing abnormal returns, I found stock prices experience much larger variation in some months than others for both an earnings beat and earnings miss. Stock prices experience the largest volatility from an earnings beat announced in June, resulting in an average 0.0035 percentage point (700% [30]) increase in daily abnormal returns. Stock prices experience the largest volatility from an earnings miss announced in December, with an on average 0.0022 percentage point (440% [31]) decrease in daily abnormal returns.

Previous literature surrounding market efficiency has thoroughly examined the role of earnings surprises on abnormal returns; however, I make contributions to this area of research in two ways. Since stock returns series generally exhibit time-varying volatility, a GARCH model can more efficiently estimate the variation in returns and thus abnormal returns by isolating the price movements that occur at the tail end of the distribution. There are very few papers which calculate abnormal returns in this manner. I contribute to these papers which provide evidence that a GARCH model can more precisely identify whether a firm is experiencing an abnormal return versus an increase in expected returns. I also add to existing papers by providing evidence investor's reactions to earnings surprises are not only asymmetric but seasonal, leading to firm's accumulation of seasonal abnormal returns. By identifying which months experience large variation in stock prices due to announcements, investors can incorporate this seasonality into their investment strategy. Financial derivatives such as options provide a source of insurance during months that suffer large variation in

[28] This is equivalent to a 340% increase in abnormal returns at the mean.
[29] This is equivalent to a 120% decrease in abnormal returns at the mean.
[30] This is equivalent to a 700% increase in abnormal returns at the mean.
[31] This is equivalent to a 440% increase in abnormal returns at the mean.

prices.

The remainder of the paper is organized as follows: Section 2 summarizes the data. Section 3 describes calculating abnormal returns and presents the methods and results for analyzing the effect earnings surprises have on abnormal returns. The last section concludes.

3.2 Data

This study uses daily stock data from the Center for Research in Security Prices (CRSP)[32] and quarterly earnings announcement data from the Institutional Brokers Estimate System (I/B/E/S). The sample period is from January 1990 to December 2019. Using daily data over the course of 29 years, I have an extremely large sample size of over 8 million observations. Since we are analyzing such rare events, it is important to have a very large data set in order to detect an effect. I drop all firms who do not have regularly scheduled quarterly earnings announcements by removing firms whose average distance between earnings announcement dates (*anndats_act*) is greater than 150 days. I chose 150 days to include late announcing firms but exclude firms that announce semi-annually in the analysis. I limit the sample of firms to U.S. publicly traded firms that were listed on the S&P 500 anytime between 1990 and 2019 to represent the entire financial market. Daily Fama-French factor data, specifically market risk premium, was collected from the Wharton Research Data Services (WRDS) website from 1990-2019 to control for daily market changes.

Average forecast is calculated for each unique earnings announcement date as the simple average of all analyst's forecasts as a proxy for investor's expectations of firm's end-of-quarter earnings. To ensure data quality, I require there to be at least 2 unique forecasts for each announcement, counted by unique analyst using I/B/E/S detail file (*analys*). After investigating earnings forecasts for a sample of 100 companies, I/B/E/S documents in it's Research Bibliography Sixth Edition that there exists a bias towards overestimation of actual earnings performance; which was determined to be result of both exogenous and endogenous factors. Earnings announcements are generally consistent in nature and the notion of overestimating earnings forecasts is also a generally consistent trend. Therefore, while it may be likely that market participants are also aware of this overestimation bias in forecasts for end-of-quarter earnings, I do not adjust forecasts for this bias since this bias comes directly from the firms. It has been shown that median earnings surprise has overtime shifted from small negative (miss analyst estimates by a small amount) to zero (meet analyst estimates exactly) to small positive (beat analyst estimates by a small amount) Brown (2001). More recently, it has

[32]For simplicity reasons, all firms with a price below one dollar were removed.

Table 7: **Summary Statistics**

Panel A: Fundamental Variables				
	Number of Days Without	Number of Days With		
Earnings Announcement	2,059,883	27,266		
Earnings Announcement w/ 3 day window	1,919,502	167,647		
Earnings Beat	2,077,485	9,664		
Earnings Beat w/ 3 day window	2,023,186	63,963		
Earnings Miss	2,079,986	7,163		
Earnings Miss w/ 3 day window	2,039,047	48,102		
Panel B: Analyst Related Variables				
	Mean	SD	1st Q	3rd Q
Number of Analysts	2.00	6.55	3.00	11.00
Forecasted Earnings	26,098	3,503,808.00	0.00	0.00
Forecast Error	-30,552	4,607,071	0.00	0.00

Note: Panel A: Count of daily end of quarter earnings announcements, count of announcements which had an earnings beat and count of announcements which had an earnings miss within the 1990-2018 sample. Additionally, included are the counts using a three-day announcement window around earnings announcement, earnings beat and earnings miss dates. Panel B: Summary statistics on the number of analysts, forecast, and forecast error for earnings announcements held within 1990-2018 for firms on the S&P 500 index. Forecast is the average of all unique forecasts for a given earnings announcement. Forecast error is actual earnings minus average forecasted earnings for a given earnings announcement.

been shown there is a positive relation between analyst coverage and whether a firm meets or beats analyst forecasts, greater analyst coverage raises the pressure on managers to meet analyst earnings forecasts. Firms will manage their earnings up a cent or two above the analysts forecasts in order to achieve a beat Huang et al. (2017), Mindak et al. (2016).

I construct two dummy variables which represent an earnings beat and an earnings miss. I define an earnings beat/miss as a 1 when the average forecasted earnings are at least one standard deviation above/below a firm's actual earnings. Additionally, although earnings announcements are consistent in nature and generally known in advance by market participants, it is possible for announcements to be late, early, or cancelled therefore, "one cannot use actual announcement dates, but rather must construct a proxy for expected announcements dates" Frazzini (2007), Cohen et al. (2007). Several prior studies regarding forecasting earnings announcement dates focus mainly on a three-day window around the announcement; see Frazzini (2007), Givoly and Palmon (1982), Chambers and Penman (1984), Begley and Fischer (1998), and Cohen et al. (2007). I address the last issue by using a 3-day announcement window around an earnings beat/miss.

Table 7 shows the descriptive statistics of my main sample as well as the earnings analysts data from I/B/E/S. Panel A describes the number of days in my sample firms had an earnings announcement in addition to how many of those days are earnings beats and earnings misses. Panel B provides some descriptive statistics of the earnings data from I/B/E/S. The unit of observation is firm-earnings-announcement. "Number of analyst" is the count of unique analysts. On average, there are two analysts that make forecasts for an individual company's earnings. The mean average forecast made for firm's end-of-quarter earnings per share was 26,098. The mean average forecast error for earnings per share was -30,552. These results show there are large negative forecast error outliers. Meaning, there were a few cases where analysts forecasts were extremely optimistic in comparison to the reality of firms' earnings. This could be the case during uncertain economic times such as a recession or financial crisis.

3.3 Calculating Abnormal Returns

Abnormal returns are calculated as the actual return minus the expected return conditional on the information set, X_i:

$$AR_{i,t} = R_{i,t} - E[R_{i,t}|X_i] \tag{11}$$

Abnormal returns are essential in determining a security's risk-adjusted performance when compared to the overall market; however, the calculation of abnormal returns in traditional event studies assumes the volatility or systematic risk of a security compared to the market is homoskedastic throughout the announcement period. Since the price reaction that occurs following an earnings surprise results in an increase to the variability of returns, I utilize a CAPM-GARCH(1,1)-M model which allows for risk to vary during the announcement period. I calculate risk-adjusted expected returns using a GARCH process of the following form:

$$R_{i,t} = \phi_0 + \beta_1 R_{m,t} + \lambda\sigma_t + a_t \tag{12}$$

$$\sigma_t^2 = \alpha_0 + \alpha_1\sigma_{t-1}^2 + \beta_1\epsilon_{t-1}^2 \tag{13}$$

Where ,

- ϕ_0 is a constant
- β_1 is a measure of the sensitivity of $R_{i,t}$ on the reference market
- $R_{m,t}$ is the market return on day day t
- λ is the volatility coefficient for the mean
- σ_t is the conditional standard deviation (i.e. volatility) at time t

- a_t is the model's residual at time t
- α_0, α_1 are the parameters of the ARCH component model
- β_1 is the parameter of the GARCH component model

Many studies that analyze volatility dynamics found a simple GARCH(1,1) model provides a good first approximation to the observed temporal dependencies in daily data Anderson and Bollerslev (1998). Additional early evidence of this can be seen in Baillie and Bollerslev (1989), Bollerslev (1987), Bollerslev et al. (1986), and Hsieh (1989). For more recent evidence see: Miah (2016) and Lunde and Hansen (2005). The unconditional distribution of a GARCH process is symmetric and leptokurtic, a similar characteristic to financial market data. Several studies show returns are not normal but leptokurtic and "fat tailed" relative to a normal distribution Officer (1972); Feng and Shi (2017), Mandelbrot (1963), Fama (1965). While the GARCH model can accommodate for excess kurtosis as a result of the propagation of shocks through time (Bollerslev and Wooldridge (1992)), there still exists in most cases excess kurtosis in the standardized residuals (Calzolari et al. (2014)). A common solution for this is to assume a fatter-tailed distribution for the error term such as a student-t distribution, since a GARCH model with the true distribution can lead to more efficient results (Bollerslev (1987)). Therefore, I use a CAPM-GARCH(1,1)-M with student-t distributed errors to calculate expected returns.

The leptokurtic nature of GARCH processes follows from the persistence in conditional variance, which produces the clusters of "low volatility" and "high volatility" episodes. According to Diebold and Lopez (1995), any of the myriad economic forces, such as seasonality, that produce persistence in economic dynamics may be responsible for GARCH effects. This persistence is in the conditional second moment, therefore it is appropriate to use a GARCH model to estimate the variance of financial market returns as the GARCH model can capture this persistence. Since the volatility builds up as the announcement date approaches and then deceases when the results of the announcement are known, a GARCH model will estimate how fast the decay in volatility is.

Unlike traditional event studies which use an estimation window and testing window to calculate abnormal returns, I use the fitted values of the CAPM-GARCH(1,1)-M model as expected returns. Unexpected returns are then calculated as the actual returns minus expected returns. This method of not using an estimation window/testing window follows similarly to that of Cutler et al. (1988) "What Moves Stock Prices?" (which also followed Roll (1984)) in estimating how much variation in returns can be attributed to news. The

Table 8: Summary Statistics

	Range of Years					
	1990 to 1994	1995 to 1999	2000 to 2004	2005 to 2009	2010 to 2014	2015 to 2019
Mean Expected Returns by CAPM-GARCH(1,1)-M						
January	-0.00006	0.00052	0.00022	-0.00091	-0.00062	-0.00047
February	0.00040	0.00044	-0.00043	-0.00136	0.00013	0.00060
March	-0.00030	0.00007	0.00015	0.00033	0.00078	0.00081
April	-0.00053	0.00035	0.00051	0.00075	0.00075	0.00053
May	0.00112	0.00064	0.00017	0.00043	-0.00013	0.00008
June	-0.00107	0.00012	0.00031	-0.00070	-0.00060	-0.00017
July	0.00060	0.00006	-0.00090	-0.00045	0.00039	0.00091
August	-0.00021	-0.00117	-0.00081	0.00029	-0.00001	-0.00029
September	-0.00090	0.00036	-0.00070	-0.00122	0.00005	0.00011
October	0.00017	0.00045	0.00091	-0.00025	0.00005	0.00047
November	-0.00025	0.00093	0.00117	0.00019	0.00020	0.00093
December	0.00121	0.00114	0.00066	0.00030	0.00053	-0.00017
Mean Abnormal Returns by Actual - Expected Returns						
January	0.00179	0.00072	0.00069	0.00099	0.00058	0.00001
February	0.00120	0.00079	0.00084	0.00072	0.00040	0.00018
March	0.00101	0.00067	0.00091	0.00118	0.00072	0.00081
April	0.00041	0.00089	0.00134	0.00130	0.00065	0.00075
May	0.00055	0.00087	0.00121	0.00097	0.00031	-0.00003
June	0.00052	0.00036	0.00070	0.00071	0.00034	0.00041
July	0.00049	0.00038	0.00049	0.00075	0.00033	0.00000
August	0.00056	0.00063	0.00086	0.00085	0.00036	0.00012
September	0.00076	0.00065	0.00060	0.00057	0.00014	0.00011
October	0.00068	0.00021	0.00049	0.00032	0.00030	0.00008
November	0.00069	0.00035	0.00099	0.00074	0.00001	0.00038
December	0.00094	0.00061	0.00095	0.00133	0.00089	0.00029

Note: Average expected returns and average abnormal returns. Expected returns are calculated as by the CAPM-GARCH(1,1)-M model. Abnormal returns are calculated as actual returns minus expected returns. All numbers are expressed as a rate.

authors relate stock returns to news about macroeconomic performance by first estimating vector auto-regressive models for the following seven macroeconomic variables: the logarithm of real dividend payments on the value-weighted New York Stock Exchange portfolio, the logarithm of industrial production, the logarithm of the real money-supply (M1), the nominal long-term interest rate (measured as Moody's AAA corporate bond), the nominal short-term interest rate (measured as the three-month Treasury Bill), and the logarithm of stock market volatility. The authors then use the residuals of these models as the unexpected component of each time series and consider the explanatory power of these "news" measures in regressions explaining stock returns. They find macroeconomic news can only explain approximately one-third of the variation in returns.

Table 16 provides an average measure of "normal" activity for firms listed on the S&P 500 between 1990-2019 through expected returns, and abnormal activity through abnormal returns. Expected returns are calculated from the CAPM-GARCH(1,1)-M model specified in Equations 12 and 13. Abnormal returns are the actual returns minus expected returns. Average expected returns across firms fluctuate relatively evenly between positive and negative values throughout the months and year ranges. In contrast, average abnormal returns are almost entirely positive throughout the months and across all year ranges. This suggests big shocks within the U.S financial market result on average in positive accumulated abnormal returns for firms with large market capitalization, regardless of the sign and month of the shock.

Of course, some months result in a larger accumulation of expected returns and/or abnormal returns than others. Both November and December have the largest average expected returns during two of the six year ranges[33], suggesting firm's expected returns are on average larger towards the end of the year[34]. In contrast, average abnormal returns are accumulated larger for firms in the beginning and end of the year, but primarily towards the beginning. April and December have the largest average expected returns during two of the six year ranges[35]. More interestingly, for any year range, if average expected returns were larger towards the end of the year, abnormal returns were accumulated larger for firms towards the beginning of year and vice versa. It should additionally be noted that the standard deviation of abnormal returns is extremely large, meaning many firm's abnormal returns in the sample vary significantly from the average. This goes to show although all firms in the sample have very large market capitalization, their abnormal returns can differ significantly between each other which I believe to be linked to investor's trading behavior. A detailed view of the variation in expected returns and abnormal returns between firms can be found in the **Appendix**.

3.4 Methods & Results

3.4.1 Asymmetric Effect of Earnings Surprises on Abnormal Returns

I utilize the abnormal returns calculated in Section 3.1 and test the role earnings announcements have on investor behavior in generating these abnormal returns. I do not believe earnings announcements are responsible for the day-to-day fluctuations in stock prices; there is roughly a coin flip toss a stock price will go up or down on a given day. I do believe, however, that they play a large role in firm's accumulation of abnormal returns. To estimate whether investor's market reactions to earnings surprises are asymmetric I use the two

[33] November has the largest average expected returns from 2000-2004 and 2015-2019. December has the largest average expected returns from 1990-1994 and 1995-1999.

[34] The largest month for average expected returns during 2005-2009 was April, and March for 2010-2014.

[35] April has the largest average abnormal returns from 1995-1999 and 2000-2004. December has the largest average abnormal returns from 2005-2009 and 2010-2014. The largest month for average abnormal returns during 1990-1994 was January, and March for 2015-2019.

dummy variables described in Section 2 to indicate an earnings beat (forecasted value was below the realized value) or earnings miss (forecasted value was above the realized value) in the following fixed effects linear regression:

$$AbnormalReturns_{it} = \beta_1 mktrf_{it} + \beta_2 ES_{it} + \beta_3 EAB_{it} + \beta_4 EAM_{it} + \boldsymbol{\alpha}_{firm} + \boldsymbol{\theta}_{year} + \epsilon_{it} \quad (14)$$

In equation (3), $mktrf$ is the market risk premium. ES represents the earnings season, which is calculated as the number of firms announcing in a given MonthYear. I control for the earnings season to control for the large variation that occurs in certain months of each quarter when a majority of publicly traded companies announce their quarterly earnings. EAM represents an earnings announcement miss, it is a binary variable that equals 1 if realized earnings were below the average forecasted earnings and a 0 otherwise. Similarly, EAB represents earnings announcement beat, it is a binary variable that represents a 1 if realized earnings were above the average forecasted earnings and a 0 otherwise. $\boldsymbol{\alpha}_{firm}$ is a vector of firm fixed effects (by CUSIP). I control for firm fixed effects to control for the firm specific idiosyncratic error. There may be large average daily returns for some firms and very small average daily returns for others. Removing the average return of each firm eliminates this issue and allows us to measure deviation from the average. $\boldsymbol{\theta}_{year}$ is a vector of year fixed effects. It is not entirely clear if abnormal returns increase from year to year as prices/returns do. Therefore, I provide results both including and excluding year fixed effects for additional robustness but utilize specification (1) for my analysis. Standard errors are clustered at the firm level.

Table 9 provides evidence concerning the impact of investor's asymmetric behavior on abnormal returns. It compares the market reaction an earnings beat has on abnormal returns with the market reaction an earnings miss produces on abnormal returns. Although my calculation of abnormal returns differs, my results are consistent with those found by Ding et al. (2004); earnings beats or positive earnings surprises are associated with significant increases in abnormal returns but earnings misses or negative earnings surprises have only a limited negative impact on returns. An earnings beat, on average, leads to a 0.0017 percentage point increase in daily abnormal returns. The average daily abnormal return is 0.0005%. This is equivalent to a 340% increase in abnormal returns at the mean. This effect is statistically significant at the 1% significance level. In contrast, an earnings miss, on average, leads to a 0.0006 percentage point decrease in daily abnormal returns or equivalently, a 120% decrease in abnormal returns at the mean. This effect is also statistically significant

Table 9: Regression Output

	Dependent variable: Abnormal Returns	
	(1)	(2)
Earnings Beat	0.00167***	0.00165***
	(0.00008)	(0.00008)
Earnings Miss	−0.00060***	−0.00072***
	(0.00019)	(0.00019)
mktrf	0.06918***	0.06966***
	(0.00362)	(0.00362)
Earnings Season	0.00000***	0.00000***
	(0.00000)	(0.00000)
Mean of Dependent Variable	0.0005	0.0005
Year FE		X
Firm FE	X	X
Observations	8,243,519	8,243,519
R^2	0.00097	0.00122
Adjusted R^2	0.00070	0.00094
F Statistic	1,997.11 (df = 8,241,281)	304.928 (df = 8,241,281)

Note: This table summarizes the average effects for earnings surprises. Earnings Beat represents a 1 if there was an earnings beat on a given date, 0 otherwise. Earnings Miss represents a 1 if there was an earnings miss on a given date, 0 otherwise. Specification (1) reports average effects without year fixed effects included. Specification (2) reports average effects with year fixed effects included. I utilize specification (1) for my analysis. Standard errors are clustered at the firm level. *p<0.1; **p<0.05; ***p<0.01

at the 1% significance level. While this average effect is still large, it is much smaller to that of an earnings beat. Kahneman and Tversky (1979) prospect theory demonstrates the dis-utility of a loss is much greater than the utility of a gain of the same magnitude. Kahneman and Tversky (1991) found that investors suffer a much greater dis-utility during a loss and are reluctant to realize their losses during negative earnings surprise. While this effect holds true, I believe the deeper effects of investor behavior from earnings surprises on abnormal returns lie within the timing of announcements. In the next section I take this analysis a step further by seasonally decomposing abnormal returns to test whether investors react more strongly to positive or negative earnings surprises dependent on the month of the announcement. Abnormal returns are essential in determining a security's or portfolio's risk-adjusted performance when compared to the overall market or a benchmark index. By identifying which months experience large variation in stock prices, investors may be able to effectively reduce risk on their position.

For an earnings beat, it may be difficult to understand the economic significance of a 0.0017 percentage point change in daily abnormal returns at first, but a simple example can show the power of this small change. For example, we can think about what would have happened to Apple's stock price if they hypothetically had an earnings beat on January 3, 2019. Apple's closing stock price was \$142.19 on that day and had a market capitalization of approximately \$611 billion (\$142.19 x 4.3 billion (shares outstanding)). Since an earning beat on average causes daily abnormal returns to increase by 0.0017 percentage points, an earnings beat would cause Apple's stock price to be 0.0017*$Price_{t-1}$ or 0.268 higher than their closing price was on January 3, 2019 without an earnings beat. In terms of company value this is equivalent to a market capitalization of approximately \$612.5 billion (\$142.46 x 4.3 billion); a 1.5 billion dollar increase in value from having an earnings beat in a single day.

3.4.2 Seasonal Effects of Earnings Surprises on Abnormal Returns

After providing evidence investor's reactions to earnings surprises are asymmetric, I additionally test whether their reactions are seasonal. That is, are investor's reactions dependent on the month of the announcement, leading to larger variation in stock prices during some months more than others. There is a reason to suspect investor's reactions to be seasonal given that negative earnings surprises have limited impact on firms' abnormal returns. To test this effect, I interact both an earnings beat (EAB) and earnings miss (EAM) with month fixed effects in the following fixed effects linear regression:

$$AbnormalReturns_{it} = \beta_1 mktrf_{it} + \beta_2 ES_{it} + \beta_3 EAB * Month_{it} + \beta_4 EAM * Month_{it} + \boldsymbol{\alpha}_{firm} + \epsilon_{it} \tag{15}$$

I exclude the main effects of an earnings beat and earnings miss in order to observe the effect of an earnings surprise in all 12 months on abnormal returns without violating perfect multicollinearity. If I were to include the main effects, I would have to drop a month to not violate perfect multicollinearity and it is not so clear as to which month should be dropped to be used as the comparison. It would be necessary to run the model 12 times, excluding each month at a time and comparing all the effects. To avoid this, I use only the interaction terms from an earnings miss and beat with month fixed effects. I make note that by including all the months, I am not able to control for the average monthly effect. Similarly to before, I provide results including and excluding year fixed effects but use specification (1) as my final model as shown in Equation 15 which excludes year fixed effects.

Table 10 provides evidence concerning the impact of investors' seasonal reactions to earnings surprises. I compare the average effect an earnings beat and earnings miss has on abnormal returns in each month. Investors' reaction to an earnings beat are extremely large when announced by firms in June. An earnings beat in June, on average, leads to a 0.0035 percentage point increase in daily abnormal returns, or equivalently, a 700% increase in abnormal returns at the mean. This effect is statistically significant at the 1% significance level. In regard to a negative surprise, an earnings miss in December decreases abnormal returns on average by 0.0022 percentage points. A lot more variation occurs in the month of December, as indicated by the higher standard error relative to other months; therefore, we cannot conclude this estimation is significant different from zero. Earnings misses announced in May, however, have a very similar effect as they do in December, with abnormal returns decreasing on average by 0.0022 percentage points. At the mean, the effect of an earnings miss announced in May is equivalent to a 440% decrease in abnormal returns. We can see again here that although a negative earnings surprise results in a decrease in firms' accumulation of abnormal returns, it is a limited downside effect across most months. Another possible explanation for this is that generally investors view the information gain that comes from an earnings announcement more positively than the actual negative earnings news.

Using a similar example to the one described in Section 3.2, it is possible to see the extreme differences in the effect an earnings beat can have on abnormal returns depending on the month of the earnings announcement. For example, I can compare what would have happened to Apple's stock price if they had an earnings beat on a random day in June, the month stock prices suffer the highest variation from earnings beats. We can then compare this scenario to what would have happened to Apple's stock price if they had an earnings beat on a random day in August, a month stock prices suffer extremely low variation from earnings beats. On June 4, 2019, Apple's closing stock price was $179.64 and had a market capitalization of approximately $826.54 billion ($179.64 x 4.6 billion (shares outstanding)). Since an earning beat in June on average causes abnormal returns to increase by 0.0035 percentage points, an earnings beat would cause Apple's stock price to be 0.0035*$Price_{t-1}$ or 0.61 higher than their closing price was on June 4, 2019 without an earnings beat. In terms of value this is equivalent to a market capitalization of $829.15 billion ($180.25 x 4.6 billion); an approximately 2.61 billion dollar increase in daily value from having an earnings beat. In comparison, an earning beat in August on average causes daily abnormal returns to increase

Table 10: Regression Output

		Dependent variable: Abnormal Returns	
		(1)	(2)
Earnings Beat	Jan	0.00142***	0.00145***
		(0.00058)	(0.00058)
	Feb	0.00254***	0.00253***
		(0.00041)	(0.00041)
	Mar	0.00165**	0.00177***
		(0.00085)	(0.00085)
	Apr	0.00190***	0.00184***
		(0.00043)	(0.00043)
	May	0.00127***	0.00130***
		(0.00051)	(0.00051)
	Jun	0.00352***	0.00363***
		(0.00141)	(0.00141)
	Jul	0.00171***	0.00163***
		(0.00052)	(0.00052)
	Aug	0.00153***	0.00155***
		(0.00053)	(0.00053)
	Sep	0.00290***	0.00295***
		(0.00114)	(0.00113)
	Oct	0.00073	0.00060
		(0.00047)	(0.00047)
	Nov	0.00128***	0.00126***
		(0.00061)	(0.00061)
	Dec	0.00266**	0.00272***
		(0.00139)	(0.00138)
Earnings Miss	Jan	−0.00054	−0.00068
		(0.00024)	(0.00024)
	Feb	−0.00092***	−0.00104***
		(0.00018)	(0.00018)
	Mar	−0.00150	−0.00152
		(0.00037)	(0.00037)
	Apr	0.00041	0.00023
		(0.00020)	(0.00020)
	May	−0.00216***	−0.00218***
		(0.00021)	(0.00021)
	Jun	0.00193	0.00194
		(0.00051)	(0.00051)
	Jul	0.00039	0.00022
		(0.00020)	(0.00020)
	Aug	−0.00104**	−0.00107**
		(0.00024)	(0.00024)
	Sep	−0.00136	−0.00132
		(0.00055)	(0.00055)
	Oct	−0.00075	−0.00097**
		(0.00021)	(0.00021)
	Nov	−0.00051	−0.00052
		(0.00026)	(0.00026)
	Dec	−0.00221	−0.00223
		(0.00060)	(0.00060)
	mktrf	0.06916***	0.06964***
		(0.00362)	(0.00362)
	Earnings Season	0.00000***	0.00000***
		(0.00000)	(0.00000)
Mean of Dependent Variable		0.0005	0.0005
Year FE			X
Firm FE		X	X
Observations		8,243,519	8,243,519
R^2		0.00100	0.00123
Adjusted R^2		0.00073	0.00095
F Statistic		330.983 (df = 8,241,230)	184.044 (df = 8,241,230)

Note: This table summarizes the average effects for earnings surprises by month. The top portion above the center line are the average effects for an earnings beat by month. The bottom portion below the center line are the average effects for an earnings miss by month. Standard Errors are clustered at the firm level. Specification (1) reports average effects without year fixed effects included. Specification (2) reports average effects with year fixed effects included. I utilize specification (1) for my analysis. Standard errors are clustered at the firm level. *p<0.1; **p<0.05; ***p<0.01

by 0.0007 percentage points. Apple had a closing price of $204.02 and market capitalization of $922 billion ($204.02 x 4.5 billion (shares outstanding)) on August 2, 2019. Therefore, an earnings beat on that day would have caused Apple's stock price to be 0.0007*$Price_{t-1}$ or 0.146 higher than their closing price was on August 2, 2019 without an earnings beat. Apple's market capitalization with an earnings beat on this day would have been $922.68 billion ($204.17 x 4.6 billion). This is approximately a 0.68 billion dollar increase in Apple's daily value; a much smaller increase compared to earnings beats released in a month with high variation. A similar example can also be down for Apple with a negative earnings surprise.

4 Conclusion

Earnings announcements have an important effect when a firm experiences an abnormal stock price movement, largely driven by investor's behavior to earnings surprises. I provide evidence that investors' reactions to asymmetric; investors react stronger on average to positive earnings surprises than negative earnings surprises. Firms experience on average a 0.0017 percentage point increase in accumulated abnormal returns from having an earnings beat, whereas an earnings miss has a limited impact on firms with on average a 0.0006 percentage point decrease in accumulated abnormal returns. One possible explanation for this is that investors are reluctant to realize their losses.

I additionally provide evidence that investor's reactions to earnings announcements are seasonal; there exists larger variation in stock prices in some months than others. By seasonally decomposing abnormal returns, I show investor's reactions to earnings announcements extremely differ in magnitude throughout the months of the year. I found investors on average have the largest reaction to an earnings beat in June and to an earnings miss in December. On average, an earnings beat in June increases daily abnormal returns by 0.0035 percentage points. An earnings miss in December decreases abnormal returns on average by 0.0022 percentage points.

Identifying which months stock prices experience larger variation from earnings surprises has some implications for the investment strategies of investors. During these high volatility months, investors can utilize financial derivatives such as options which provide insurance on their position. For example, a call option allows an investor to sell their stock at a specific price by a given date. The call option would allow an investor to sell their stock at a previously agreed upon price regardless if the stock price drops during one of these highly volatile months. This means the investor is protected from any downside risk during the time the option is effective. It is well known earnings season there exists a lot of volatility; many stocks exceed their earnings estimates and experience a big jump in price, and several others fall short of their estimates and sustain a big price drop. However, it should be noted higher volatility means higher option prices. My findings allow for investors to identify which months are worth purchasing that additional insurance on their position than others.

References

Aıt-Sahalia, Y., Wang, Y., and Yared, F. (2001). Do option markets correctly price the probabilities of movement of the underlying asset? *Journal of Econometrics*, 102(1):67–110.

Anderson, T. G. and Bollerslev, T. (1998). Modeling volatility dynamics. *International Economic Review*, 39(4).

Ang, A., Bolvin, J., Dong, S., and Loo-Kung, R. (2011). Monetary policy shifts and the term structure. *The Review of Economic Study*, 78(2):429–457.

Ang, A. and Piazzesi, M. (2003). A no-arbitrage vector autoregression of term structure dynamics with macroeconomic and latent variables. *Monetary Economics*, 50:745–787.

Baillie, R. T. and Bollerslev, T. (1989). Capital market seasonality: The case of stock returns. *American Statistical Association*, 7(3):297–305.

Ball, R. and Brown, P. R. (1969 and 2014). A retrospective. *The Accounting Review*, 89(1):1–26.

Ball, R. and Kothari, S. P. (1991). Security returns around earnings announcements. *The Accounting Review*, 66(4):718–738.

Bates, D. S. (2000). Post-'87 crash fears in the sp 500 futures option market. *Journal of Econometrics*, (94):181–238.

Beaver, W. H., McNichols, M. F., and Wang, Z. Z. (2017). Increased information content of earnings announcements in the 21st century: An empirical investigation. *Stanford Working Paper No. 3616*.

Begley, J. and Fischer, P. E. (1998). Timeliness of reporting and the stock price reaction to earnings announcements. *Review of Accounting Studies*, 3(4):347–363.

Black, F. and Scholes, M. (1973). The pricing of options and corporate liabilities. *Journal of Political Economy*, 81(3):637–654.

Bollerslev, T. (1987). A conditionally heteroskedastic time series model for speculative prices and rates of return. *The Review of Economics and Statistics*, 69(3):542–547.

Bollerslev, T., Engle, R. F., and Nelson, D. B. (1986). Arch models. *Handbook of Econometrics*, 4:2959–3038.

Bollerslev, T. and Wooldridge, J. M. (1992). Quasi maximum likelihood estimation and inference in dynamic models with time varying covariances. *Econometric Reviews*, 11(2):143–172.

Bomfim, A. N. (2003). Monetary policy and the yield curve. *Board of Governors of the Federal Reserve System (U.S.)*, Finance and Economics Discussion Series 2003-15.

Brand, C., Turunen, J., and Buncic, D. (2006). The impact of ecb monetary policy decisions and communication on the yield curve. *European Central Bank*, Working Paper Series 657.

Brown, L. D. (2001). A temporal analysis of earnings surprises: Profits versus losses. *Journal of Accounting Research, Wiley Blackwell*, 39(2):221–241.

Calzolari, G., Halbleib, R., and Parrini, A. (2014). Estimating garch-type models with symmetric stable innovations: Indirect inference versus maximum likelihood. *Computational Statistics Data Analysis*, 76:158–171.

Cavallo, A. and Rigobon, R. (2016). The billion prices project: Using online prices for inflation measurement and research. *Journal of Economic Perspectives*, 30(2):151–178.

Chambers, A. E. and Penman, S. H. (1984). Timeliness of reporting and the stock price reaction to earnings announcements. *Journal of Accounting Research*, 22(1):21–47.

Choi, H. and Varian, H. R. (2009). Predicting the present with google trends. *Available at SSRN: https://ssrn.com/abstract=1659302 or http://dx.doi.org/10.2139/ssrn.1659302*.

Cohen, D. A., Dey, A., Lys, T. Z., and Sunder, S. V. (2007). Earnings announcement premia and the limits to arbitrage. *Journal of Accounting and Economics*, 43(2-3):153–180.

Cutler, Poterba, and Summers (1988). What moves stock prices? *NBER Working Paper Series*, (2538).

Diebold, F. and Rudebusch, G. (2013). Yield curve modeling and forecasting: The dynamic nelson-siegel approach. (yield curve modeling and forecasting.). *Princeton: Princeton University Press*.

Diebold, F., Rudebusch, G., and Aruoba, S. (2004). The macroeconomy and the yield curve: A dynamic latent factor approach. *National Bureau of Economic Research, Inc*, NBER Working Papers 10616.

Diebold, F. X. and Lopez, J. A. (1995). Modeling volatility dynamics. *NBER*, (173).

Diebold, F. X., Tay, A. S., and Wallis, K. F. (1997). Evaluating density forecasts of inflation: The survey of professional forecasters. *National Bureau of Economic Research, Inc*, NBER Working Papers 6228.

Ding, D. K., Charoenwong, C., and Seetoh, R. (2004). Prospect theory, analyst forecasts, and stock returns. *Journal of Multinational Financial Management*, 14(4–5):425–442.

Duffee, G. (2013). Forecasting interest rates," handbook of economic forecasting. *Elsevier*.

Engelberg, J. R., McLean, D., and Pontiff, J. (2018). Anomalies and news. *Journal of Finance*, 73(5):1971–2001.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica*, 50(4):987–1007.

Evans, C. L. and Marshall, D. A. (2007). Economic determinants of the nominal treasury yield curve. *Journal of Monetary Economics, Elsevier*, 54(7):1986–2003.

Fama, E. (1969). Efficient capital markets: A review of theory and empirical work. *American Finance Association*, 25(2):28–30.

Fama, E. F. (1965). Portfolio analysis in a stable paretian market. *Management Science*, 11(3):404–419.

Feng, L. and Shi, Y. (2017). A simulation study on the distributions of disturbances in the garch model. *Cogent Economics Finance*, 5(1):1355503–135.

Fleming, J. (1998). The quality of market volatility forecasts implied by s&p 100 index option prices. *Journal of Empirical Finance*, 5(4):317–345.

Francis, J., Schipper, K., and Vincent, L. (2002). Expanded disclosures and the increased usefulness of earnings announcements. *The Accounting Review*, 77(3):515–546.

Frazzini, L. (2007). The earnings announcement premium and trading volume. *NBER Working Paper Series*, (13090).

Geske, R. and Roll, R. (1984). On valuing american call options with the black-scholes european formula. *The Journal of Finance*, 39(2):443–455.

Givoly, D. and Palmon, D. (1982). Earnings announcement premia and the limits to arbitrage. *The Accounting Review*, 57(3):486–508.

Haldane, A. G. and Read, V. (2000). Monetary policy surprises and the yield curve. *Bank of England*, Bank of England working papers 106.

Hsieh, D. A. (1989). Modeling heteroscedasticity in daily foreign-exchange rates. *Journal of Business Economic Statistics*, 7(3):307–317.

Huang, S. X., Pereira, R., and Wang, C. (2017). Analyst coverage and the likelihood of meeting or beating analyst earnings forecasts. *Contemporary Accounting Research, John Wiley - Sons*, 34(2):871–899.

Kahneman, D. and Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, 47(2):263–291.

Kahneman, D. and Tversky, A. (1991). Loss aversion in riskless choice: a reference-dependent model. *Quarterly Journal of Economics*, 106:1039–1061.

Kownatzki, C. (2016). How good is the vix as a predictor of market risk? *Journal of Accounting and Finance*, 16(6).

Kownatzki, C. and Sabouni, H. (2019). Option strangles: An analysis of selling equity insurance. *Available at SSRN 3315014*.

Kuttner, K. N. (2000). Monetary policy surprises and interest rates: evidence from the fed funds futures markets. *Federal Reserve Bank of New York*, Staff Reports 99.

Landsman, W. R. and Maydew, E. L. (2002). Has the information content of quarterly earnings announcements declined in the past three decades? *Journal of Accounting Research*, 40(3):797–808.

Lehmann, E. L. and Romano, J. P. (2006). *Testing Statistical Hypotheses*. Springer Science & Business Media.

Linnainmaa, J. T. and Zhang, Y. (2019). The earnings announcement return cycle. *Available at SSRN: https://ssrn.com/abstract=3183318 or http://dx.doi.org/10.2139/ssrn.3183318.*

Lunde, A. and Hansen, P. R. (2005). A forecast comparison of volatility models: does anything beat a garch(1,1)? *Journal of Applied Econometrics*, 20(7):873–889.

Mandelbrot, B. (1963). The variation of certain speculative prices. *The Journal of Business*, 36(4):394–419.

Merton, R. C. (1973). Theory of rational option pricing. *The Bell Journal of Economics and Management Science*, pages 141–183.

Miah, M. (2016). Modelling volatility of daily stock returns: Is garch(1,1) enough? *American Scientific Research Journal for Engineering, Technology, and Sciences*, 18(1):29–39.

Mindak, M. P., Sen, P. K., and Stephan, J. (2016). Beating threshold targets with earnings management. *Review of Accounting and Finance, Emerald Group Publishing*, 15(2):198–221.

Mumtaz, H. and Surico, P. (2008). Time-varying yield curve dynamics and monetary policy. *Monetary Policy Committee Unit, Bank of England*, Discussion Papers 23.

Officer, R. R. (1972). The distribution of stock returns. *Journal of the American Statistical Association*, 67(340):807–812.

Poon, S.-H. and Granger, C. W. (2003). Forecasting volatility in financial markets: A review. *Journal of Economic Literature*, 41(2):478–539.

Roll, R. (1984). Orange juice and weather. *The American Economic Review*, 74(5):861–880.

Rubinstein, M. (1994). Implied binomial trees. *Journal of Finance*, (49):771–818.

Sabouni, H. (2018). The rhythm of markets. *Claremont Graduate University.*

Savor, Pavel, and Wilson (2016). Earnings announcements and systematic risk. *Journal of Finance*, 71(13090):83–138.

Scholz, F. W. and Stephens, M. A. (1987). K-sample anderson-darling tests. *Journal of the American Statistical Association*, 82(399):918–924.

Taylor, J. B. (1993a). Discretion versus policy rules in practice. *Carnegie-Rochester Series on Public Policy, North-Holland*, 39:195–214.

Thaler, R. H. (1987). Anomalies: The january effect. *The American Economic Review*, 1(1):197–201.

Tsatsaronis, K. and Smets, F. (1997). Why does the yield curve predict economic activity? dissecting the evidence for germany and the united states. *Bank for International Settlements*, BIS Working Papers 49.

Wood, J. H. (1964). The expectations hypothesis, the yield curve, and monetary policy. *The Quarterly Journal of Economics, Oxford University Press*, 78(3):457–470.

5 Appendix

American versus European style Options

We find the need to distinguish American and European style options. Our paper utilizes options chains on SPY, which is traded American style. The BSM assumptions are meant for pricing and characterizing the underlying asset in the form of European style options, where the holder of the option can exercise only at expiration. For American style options, the holder can exercise at an point in time. This presents a contrasting difference in pricing an option, where American style options are greater than or equal to the price of European. However, the BSM can approximate American options fairly close to European. The same underlying assumptions regarding the volatility smile and distributional assumptions still hold in this analysis by implementing the BSM on SPY. Our focus is to estimate parameters in the model, specifically $N(d_2)$ from the original BSM framework, rather than observing price differences, which may lead to biased results, from American and European call options Geske and Roll (1984).

Implied Volatility from Dataset

Our analysis utilizes data from IVolatility.com on SPY options contracts, which provides estimates on the BSM inputs of implied volatility, time to maturity, option price, strike, and spot price (or closing price of SPY). We clarify here the use of implied volatility, which is provided by our data source for every strike in an options chain. IVolatility applies the Cox, Ross and Rubenstein method (CRR) to estimate implied volatility of SPY using a binomial tree of 100 steps. The method of CRR is applied primarily due to early exercise ability of SPY options, and dividend component. Dividends expected to paid on the underlying asset cause the price to drop, and may affect the early exercise of the option. To adjust for the early exercise component, the CRR approach provides a better estimate of implied volatility than the BSM framework.

Implied BSM versus Implied Monte Carlo

In addition to testing the forecasting accuracy of BSM, we run a Monte Carlo simulation with 1,000,000 iterations to observe a simulated path of risk adjusted probabilities. The simulated Monet Carlo options pricing model is the same as Black-Scholes with the exception that one measure of volatility is used (average of past 180-days) in comparison to BSM which uses the forecasted volatility (implied volatility). Figure 7 displays the implied CDF from a Monte Carlo simulation for all unique option expiration dates 180 days to expiration. We simulate the price path of SPY using the market price at 180 days prior to expiration, and applying an analytical solution of a Geometric Brownian motion in the form of:

$$S_t = S_0 e^{(\bar{r_f} - d - \frac{1}{2}\sigma^2)t + \sigma z \sqrt{t}} \tag{16}$$

Here, $\bar{r_f}$ is the risk free rate, d is the dividend yield of SPY, σ is the constant volatility, z is random normal variable with mean zero and standard deviation of one, and t is the time step of the price path. We then bin the simulated price path using strike prices in the options chain, and create an empirical distribution of risk adjusted probabilities. We find that the Monte Carlo simulation results in a similar time path and shift of risk adjusted probabilities as we see in Figure 2. Over time and closer to expiration, there is less probability of deep out-of-the money options being exercised. Such an observation suggests that high uncertainty about the overall price path of SPY gives way to a more uniform distribution of $N(d_2)$ closer to maturity. This is consistent with the results we saw in Figure 2; much of the uncertainty can be attributed to implied volatility from the BSM overestimating the risk attributed to deep in and out-the money contracts. The probability mass of the tails decreases as an options chain approaches expiration, suggesting traders have a much more clearer picture as to which strikes are likely to remain in-the-money at expiration.

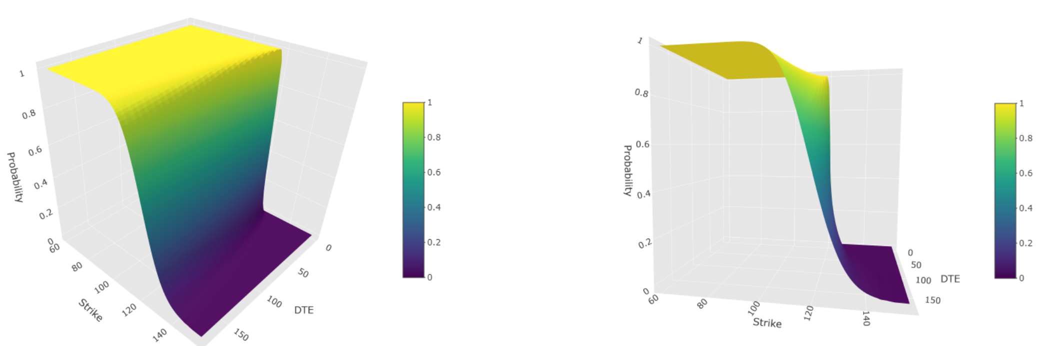


Figure 7: **Monte Carlo Simulation of Risk Adjusted Probabilities**

Note: A Monte Carlo simulation applied on 1,000,000 iterations for 180 days to expiration. The plot uses the same expiration dates as that of Figure 2 , and simulates the price path of SPY 180 days to expiration. The price path is then binned by the same strike prices given in Figure 2 for a call option. A more detailed look of the Monte Carlo simulated implied cumulative distribution can be seen.

Observing Figure 7, we can see how well the simulated Monte Carlo model predicts the underlying price path of SPY furthest from expiration. The simulated Monte Carlo model, in comparison to BSM, assigns less weight to the tails of the distribution; resulting in a more S-shaped curve farther out from expiration. Implied volatility for deep in and out of the-money accounts for this difference, since we assume a constant volatility for the Monte Carlo simulation. With a greater number of iterations, the resulting Monte Carlo simulation converges to the implied distribution of the BSM in Figure 2.

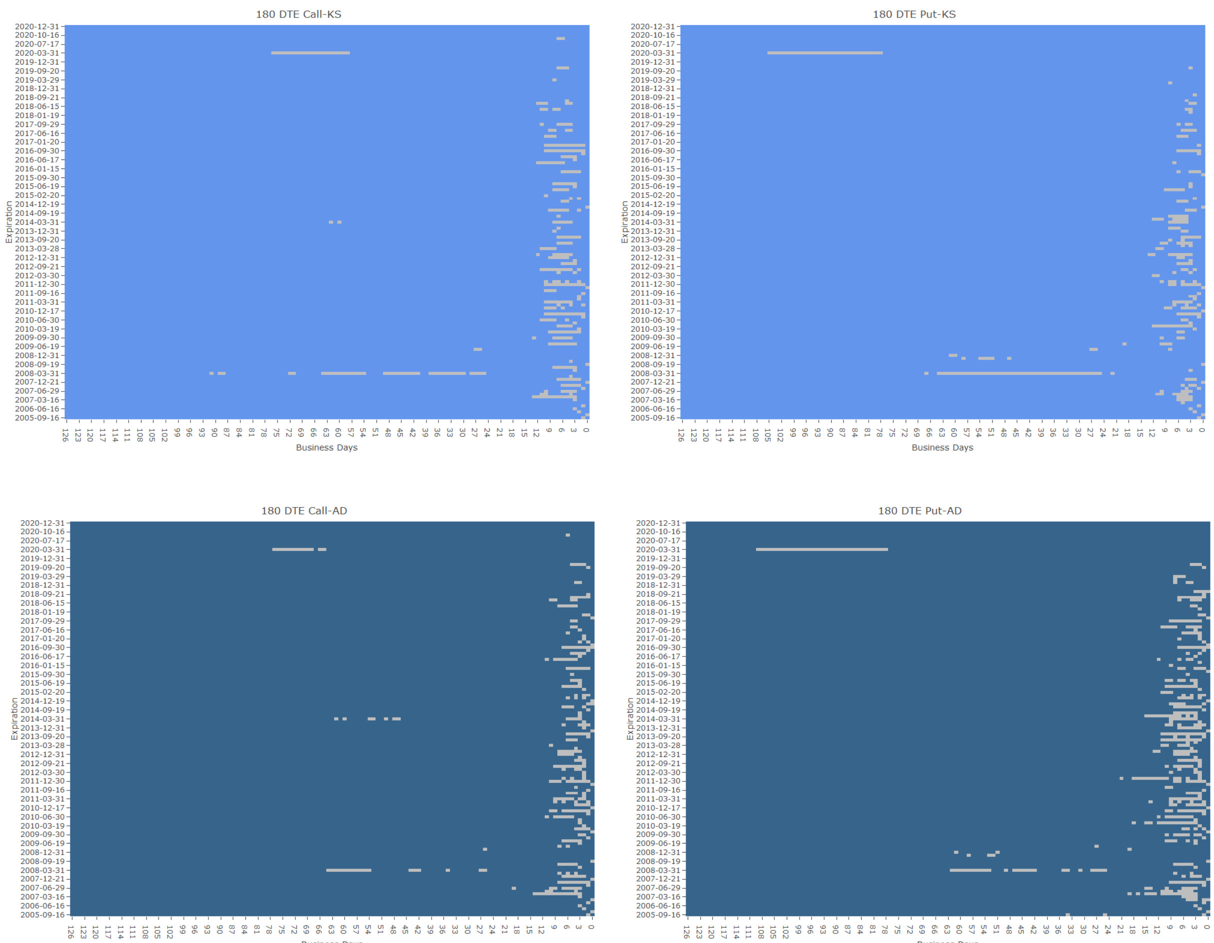


Figure 8: In Depth Summary Results for 180 DTE

Note: Implied BSM distribution and actual distribution tested under the K-S (top two sub-figures) and A-D (bottom two sub-figures) test for each unique expiration date across 180 days to expiration. The blue and red colors mean the null hypothesis was rejected on that day, meaning the two distributions do not come from the same underlying population distribution. The green and yellow colors mean we failed to reject the null hypothesis on a given day, meaning the two distributions are from the same underlying population distribution.

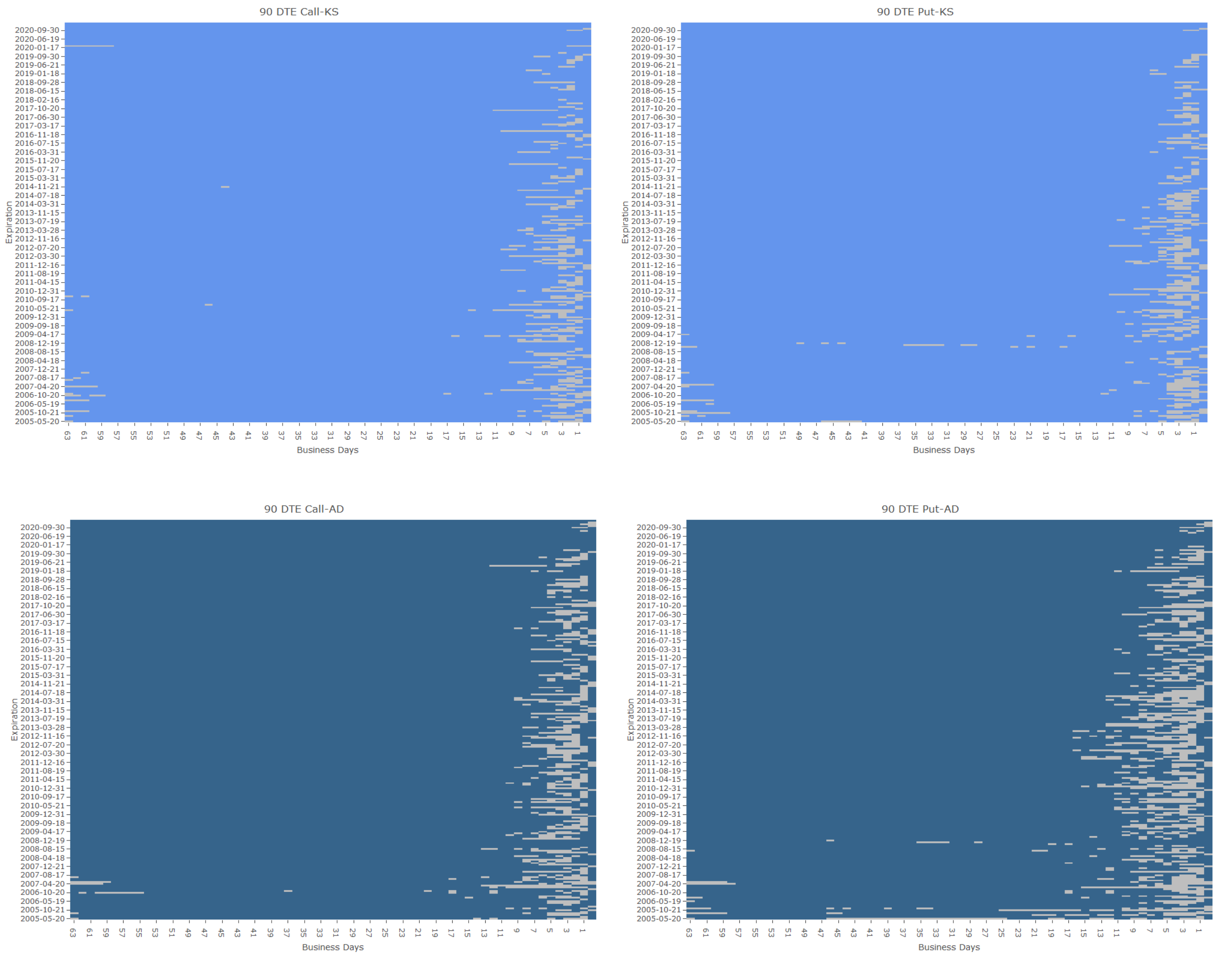


Figure 9: In Depth Summary Results for 90 DTE

Note: Implied BSM distribution and actual distribution tested under the K-S (top two sub-figures) and A-D (bottom two sub-figures) test for each unique expiration date across 90 days to expiration. The blue and red colors mean the null hypothesis was rejected on that day, meaning the two distributions do not come from the same underlying population distribution. The green and yellow colors mean we failed to reject the null hypothesis on a given day, meaning the two distributions are from the same underlying population distribution.

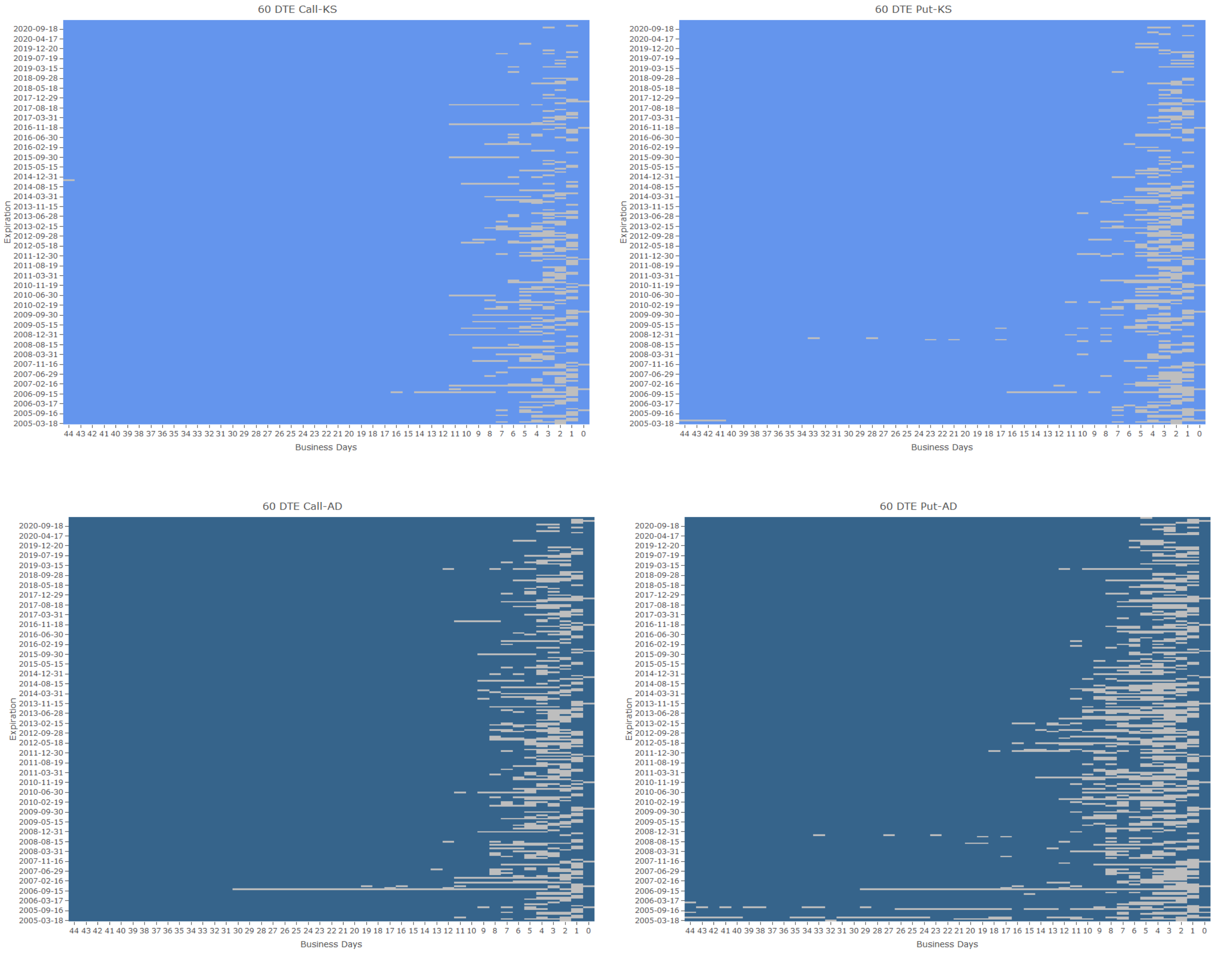


Figure 10: In Depth Summary Results for 60 DTE

Note: Implied BSM distribution and actual distribution tested under the K-S (top two sub-figures) and A-D (bottom two sub-figures) test for each unique expiration date across 60 days to expiration. The blue and red colors mean the null hypothesis was rejected on that day, meaning the two distributions do not come from the same underlying population distribution. The green and yellow colors mean we failed to reject the null hypothesis on a given day, meaning the two distributions are from the same underlying population distribution.

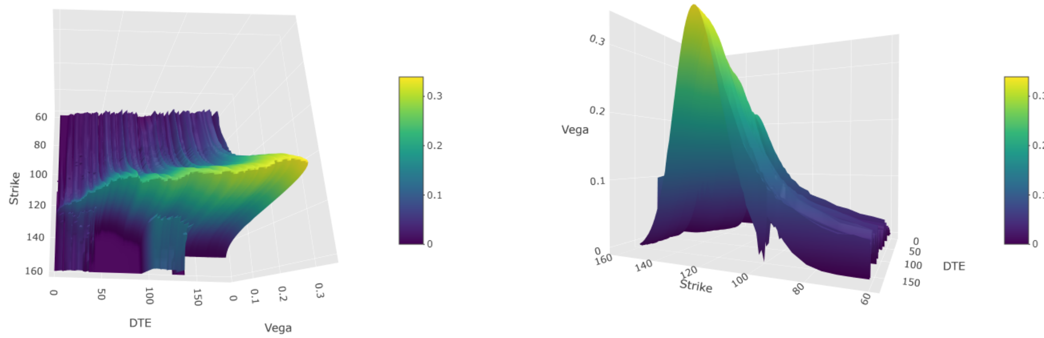


Figure 11: **Sample Vega Surface**

Note: The above plot displays a vega surface for a call options chain that expires on 2005-12-16 from our sample SPY data. Vega tends to peak ATM, and is much lower as the options approaches expiration.

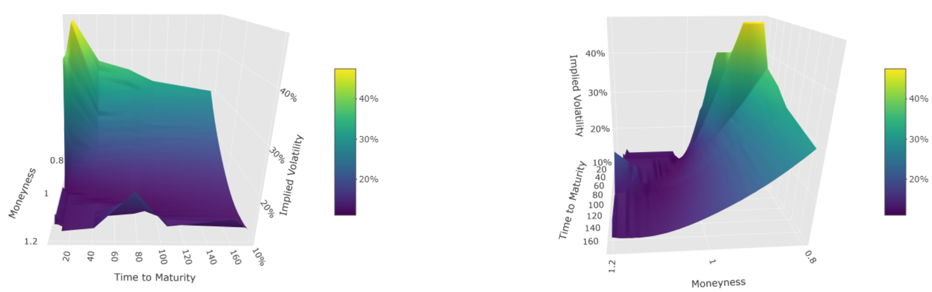


Figure 12: **Sample IV Surface for Calls**

Note: A sample implied volatility surface is displayed above, provided by a sample of data that expires on 2012-10-05. The surface incorporates the term structure and volatility smile of call options data. For out of the-money options, the implied volatility peaks. The term structure portion shows that implied volatility peaks for short dated options, as opposed to options with a longer time to maturity.

Table 10: Average Implied Volatility Differential for Call Options

	Range of Moneyness K/S_0				
Year	[,0.85)	[0.85,0.95)	[0.95,1.05)	[1.05,1.15)	[1.15,)
2005	0.2497	0.0933	0.0073	0.0109	0.0302
2006	0.3307	0.0998	0.0079	0.0019	0.0180
2007	0.3575	0.0981	0.0061	-0.0260	-0.0001
2008	0.3795	0.0856	0.0026	-0.0354	0.0033
2009	0.3195	0.0965	0.0050	-0.0261	0.1070
2010	0.2875	0.1051	0.0077	-0.0249	0.0565
2011	0.3024	0.1036	0.0087	-0.0280	0.0295
2012	0.3062	0.1096	0.0112	-0.0169	0.0715
2013	0.3029	0.1221	0.0159	-0.0026	0.0727
2014	0.3249	0.1270	0.0138	-0.0129	0.0810
2015	0.3112	0.1245	0.0081	-0.0281	0.0533
2016	0.3213	0.1264	0.0114	-0.0126	0.1130
2017	0.3949	0.1395	0.0219	0.0081	0.0914
2018	0.4336	0.1287	0.0137	-0.0057	0.0886
2019	0.4248	0.1250	0.0132	-0.0192	0.0866
2020	0.3662	0.1430	0.0116	-0.0366	-0.0298

Note: The table summarizes the mean implied volatility differential. We do so by subtracting implied volatility for each contract in a call options chain by the ATM contract (contract with moneyness closest to 1). We calculate the average based on a range of moneyness, alowing us to observe how volatility changes across options that are ITM, ATM, and OTM.

Table 11: Average Implied Volatility Differential for Put Options

	Range of Moneyness K/S_0				
Year	[,0.85)	[0.85,0.95)	[0.95,1.05)	[1.05,1.15)	[1.15,)
2005	0.2419	0.0861	0.0071	0.0357	0.1133
2006	0.2887	0.0902	0.0075	0.0286	0.1469
2007	0.2866	0.0844	0.0054	-0.0089	0.0720
2008	0.2619	0.0640	0.0026	-0.0240	0.0429
2009	0.2244	0.0728	0.0045	-0.0158	0.1290
2010	0.2092	0.0821	0.0072	-0.0132	0.0777
2011	0.2268	0.0823	0.0083	-0.0157	0.0426
2012	0.2104	0.0808	0.0099	-0.0012	0.0812
2013	0.1884	0.0960	0.0151	0.0162	0.0903
2014	0.2061	0.0986	0.0136	0.0018	0.0891
2015	0.2207	0.0930	0.0094	-0.0128	0.1015
2016	0.2232	0.1014	0.0122	0.0034	0.1516
2017	0.2871	0.1227	0.0237	0.0375	0.1171
2018	0.3380	0.1202	0.0157	0.0091	0.1389
2019	0.3485	0.1100	0.0162	0.0006	0.1368
2020	0.3450	0.1333	0.0133	-0.0361	-0.0230

Note: The table summarizes the mean implied volatility differential. We do so by subtracting implied volatility for each contract in a put options chain by the ATM contract (contract with moneyness closest to 1). We calculate the average based on a range of moneyness, alowing us to observe how volatility changes across options that are ITM, ATM, and OTM.

Table 12: Mean of Options Greeks by Moneyness Range

(a) Call Options

	Range of Moneyness K/S_0				
	[,0.85)	[0.85,0.95)	[0.95,1.05)	[1.05,1.15)	[1.15,)
delta	0.9479	0.8694	0.4818	0.0722	0.0299
vega	0.0901	0.1733	0.2210	0.1038	0.0487
gamma	0.0027	0.0100	0.0358	0.0088	0.0031
theta	-0.0240	-0.0502	-0.0953	-0.0219	-0.0161
rho	0.2543	0.2397	0.1190	0.0336	0.0122

(b) Put Options

	Range of Moneyness K/S_0				
	[,0.85)	[0.85,0.95)	[0.95,1.05)	[1.05,1.15)	[1.15,)
delta	-0.0224	-0.0992	-0.4783	-0.8950	-0.9570
vega	0.0463	0.1424	0.2281	0.1431	0.0595
gamma	0.0016	0.0087	0.0375	0.0116	0.0038
theta	-0.0134	-0.0434	-0.1081	-0.0538	-0.0466
rho	-0.0137	-0.0492	-0.1192	-0.2507	-0.3364

Note: We present summary statistics for the Greeks for call and put options separately. From 2005 to 2020, we take the mean of each options Greeks within a range of moneyness.

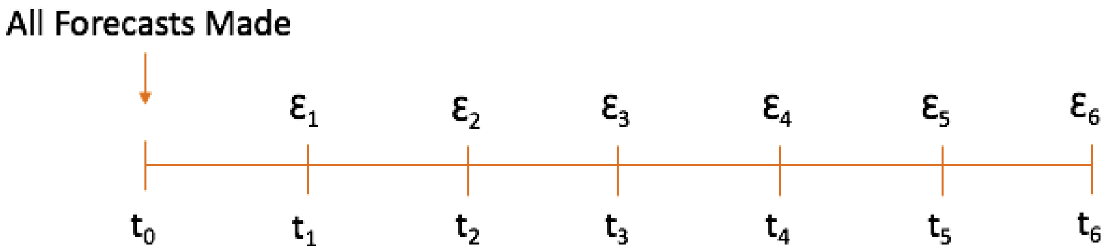


Figure 13: Timeline of SPF Forecasts

Note: SPF forecasts for one to six quarters out were made at t_0 for each security's yield. Forecast error, or ϵ, is defined as the actual yield minus the average forecasted yield at each quarterly projection.

Table 13: Descriptive Statistics for Three-Month Treasury Bill SPF Forecast Error

Three-Month Treasury Bill				
Quarters Ahead	Mean	SD	1st Q	3rd Q
1	-0.093	0.948	-0.381	0.376
2	-0.128	0.988	-0.472	0.374
3	-0.240	1.169	-0.801	0.433
4	-0.407	1.414	-1.056	0.502
5	-0.567	1.590	-1.427	0.441
6	-0.732	1.709	-1.901	0.347

Note: SPF forecast error from 1980 Q4 to 2018 Q3 for the three-month Treasury bill. Forecast error is displayed for projections made 1-6 quarters ahead of time. Forecasts are the average of all unique forecasts for a given quarter. Forecast error is actual yield minus the average forecasted yield made one quarter ahead of time. All numbers are expressed as a rate.

Table 14: Descriptive Statistics for Moody's AAA Corporate Bond SPF Forecast Error

Moody's AAA Corporate Bond				
Quarters Ahead	Mean	SD	1st Q	3rd Q
1	NA	NA	NA	NA
2	-0.192	0.600	-0.513	0.165
3	-0.303	0.741	-0.665	0.027
4	-0.433	0.866	-0.787	-0.042
5	-0.563	0.943	-0.931	-0.184
6	-0.696	0.975	-1.111	-0.203

Note: SPF forecast error from 1983 Q1 to 2018 Q3 for Moody's AAA corporate bond. Forecast error is displayed for projections made 2-6 quarters ahead of time. Forecasts are the average of all unique forecasts for a given quarter. Forecast error is actual yield minus the average forecasted yield made one quarter ahead of time. All numbers are expressed as a rate.

Table 15: Descriptive Statistics for Ten-Year Treasury Note SPF Forecast Error

Ten-Year Treasury Note				
Quarters Ahead	Mean	SD	1st Q	3rd Q
1	-0.096	0.573	-0.443	0.286
2	-0.170	0.631	-0.581	0.216
3	-0.305	0.730	-0.833	0.089
4	-0.446	0.814	-1.047	0.032
5	-0.579	0.840	-1.107	-0.128
6	-0.705	0.836	-1.267	-0.188

Note: SPF forecast error from 1980 Q4 to 2018 Q3 for the ten-year Treasury note. Forecast error is displayed for projections made 1-6 quarters ahead of time. Forecasts are the average of all unique forecasts for a given quarter. Forecast error is actual yield minus the average forecasted yield made one quarter ahead of time. All numbers are expressed as a rate.

Table 16: Summary Statistics

	Range of Years					
	1990 to 1994	1995 to 1999	2000 to 2004	2005 to 2009	2010 to 2014	2015 to 2019
SD of Expected Returns by CAPM-GARCH(1,1)-M						
January	0.00718	0.00787	0.01128	0.01336	0.01081	0.01098
February	0.00740	0.00783	0.00941	0.01149	0.01056	0.01116
March	0.00579	0.00706	0.01163	0.01525	0.01226	0.01134
April	0.00754	0.00801	0.01245	0.01325	0.00997	0.00932
May	0.00583	0.00711	0.01031	0.01100	0.01031	0.00948
June	0.00565	0.00691	0.00895	0.00999	0.01051	0.00976
July	0.00523	0.00741	0.01184	0.01251	0.00985	0.00789
August	0.00795	0.01006	0.01122	0.01055	0.01254	0.01172
September	0.00589	0.00947	0.01244	0.01487	0.01340	0.00966
October	0.00724	0.01078	0.01348	0.01959	0.01824	0.01063
November	0.00695	0.00738	0.00951	0.01655	0.01591	0.00933
December	0.00530	0.00714	0.00968	0.01316	0.01204	0.00998
SD of Abnormal Returns by Actual - Expected Returns						
January	0.02872	0.02919	0.03149	0.03053	0.02376	0.02327
February	0.02853	0.02789	0.02865	0.02778	0.02295	0.02384
March	0.02677	0.02744	0.02998	0.02987	0.02284	0.02264
April	0.02844	0.02985	0.03171	0.03002	0.02327	0.02213
May	0.02603	0.02650	0.02739	0.02943	0.02533	0.02444
June	0.02669	0.02701	0.02803	0.02537	0.01985	0.01894
July	0.02670	0.02755	0.02949	0.02818	0.02254	0.02062
August	0.02683	0.02722	0.02723	0.02558	0.02296	0.02210
September	0.02672	0.02834	0.02916	0.02689	0.02156	0.01902
October	0.02856	0.03078	0.03211	0.03223	0.02731	0.02192
November	0.02708	0.02863	0.02869	0.02922	0.02572	0.02283
December	0.03173	0.03221	0.03027	0.02800	0.02327	0.01951